

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: August 22, 2003, 01:28:08 ; Search time 3471 Seconds  
(without alignments)  
10253.925 Million cell updates/sec  
Title: US-09-831-142B-1  
Perfect score: 870  
Sequence: 1 gaattcgacagtcgaa.....aaaaaaaaaaaaactcgag 870

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2888711 seqs, 20454813386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_btg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	870	100.0	870	6	AX024884 Sequence
2	850	97.7	850	3	PDA131051
3	834.8	96.0	846	3	PDA131053
4	834.8	96.0	852	6	AX024886 Sequence
5	818	94.0	818	3	PDA131054
6	799.2	91.6	816	6	AX024885 Sequence
7	797.2	91.6	802	3	PDA131052
8	685.4	78.8	695	3	AF497971
9	52.8	6.1	1995	10	BC050759 Mus muscu
10	51.8	6.0	349980	6	AX344567 Sequence
11	51.6	5.9	7218	6	I66494 Sequence 14
12	48.2	5.5	272698	3	PFMAL4P4
13	47.4	5.4	9810	6	AX345329
14	46.8	5.4	1716	3	AEN508686
15	46.8	5.4	1716	3	AEN508687
16	46.4	5.3	659	9	BC022405
17	46.4	5.3	853	3	AY119449
18	46.2	5.3	5543	3	AY160095
19	46.2	5.3	9741	6	AX346225
20	46.2	5.3	236824	2	AC097985
21	46	5.3	151802	3	AC114263
22	46	5.3	179938	2	AC141012
23	45.8	5.3	2955	10	BC021442
24	45.8	5.3	110000	3	AC116957_1
25	45.8	5.3	125958	3	AC115592
26	45.8	5.3	205281	2	AC139217
27	45.8	5.3	219204	2	AC110090
28	45.8	5.3	226764	2	AC095754
29	45.6	5.2	3028	9	BSM801560
30	45.4	5.2	546	10	BC026533
31	45.4	5.2	1257	9	BC043335
32	45.4	5.2	1604	9	BC044260
33	45.4	5.2	2091	6	BD176856
34	45.4	5.2	2091	17	AF111849
35	45.4	5.2	2158	9	AB070046
36	45.4	5.2	136240	3	AC117070
37	45.4	5.2	178273	2	AC005308
38	45.4	5.2	230531	3	AE014845
39	45.4	5.2	267290	2	AC098041
40	45.4	5.2	303091	2	AC084799
41	45.4	5.2	349980	6	AX344553
42	45.4	5.2	349980	6	AX344554
43	45.2	5.2	432	9	BC040830
44	45.2	5.2	110000	2	PFMAL13_20
45	45.2	5.2	162871	3	AC117176

ALIGNMENTS

RESULT 1  
AX024884  
LOCUS  
DEFINITION Sequence 1 from Patent WO0028025.  
ACCESSION AX024884  
VERSION AX024884.1 GI:10184825  
KEYWORDS  
SOURCE Pholas dactylus  
ORGANISM Pholas dactylus  
Rukaryota; Metazoa; Mollusca; Bivalvia; Heteroconchia; Veneroidea;  
Pholadoidea; Pholadidae; Pholas.  
REFERENCE 1  
AUTHORS Campbell,A.K.  
TITLE Pholasin  
JOURNAL Patent: WO 0028025-A 1 18-MAY-2000;

AX024884 870 bp DNA linear PAT 15-SEP-2000

band date

UNIV WALES MEDICINE (GB); CAMPBELL ANTHONY KEITH (GB)

```
FEATURES
  source
    1. .870
      /organism="Pholas dactylus"
      /mol_type="genomic DNA"
      /db_xref="taxon:52916"
BASE COUNT      289 a   163 c   213 g   205 t
ORIGIN
Query Match      100.0%; Score 870; DB 6; Length 870;
Best Local Similarity 100.0%; Pred. No. 1.8e-208;
Matches 870; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATTCGGCAGGAGTCGGAAGACAAATGCGTTCATCGTTTCCTTCCTTCGTCG 60
Db 1 GAATTCGGCAGGAGTCGGAAGACAAATGCGTTCATCGTTTCCTTCCTTCGTCG 60
QY 61 CTCATGCTTAATGCAACCGGTTCCGGTGAGGAAGTACAATGCGCGATGAATGGACAC 120
Db 61 CTCATGCTTAATGCAACCGGTTCCGGTGAGGAAGTACAATGCGCGATGAATGGACAC 120
QY 121 AAGCTAATGATATGCTTCAACGTGGACGTGGATGACCAATTTTATCTACGACTATGGCG 180
Db 121 AAGCTAATGATATGCTTCAACGTGGACGTGGATGACCAATTTTATCTACGACTATGGCG 180
QY 181 CTCAGAGCAACTGTACGAAGATCGGGCTTTGGGGCTGTGTCGGATTGAAGCGGCCGCC 240
Db 181 CTCAGAGCAACTGTACGAAGATCGGGCTTTGGGGCTGTGTCGGATTGAAGCGGCCGCC 240
QY 241 CAGTACACAAAGCCGCTCGGATTAACGTGAGTACAGACAGCAGTCATGTGTACAA 300
Db 241 CAGTACACAAAGCCGCTCGGATTAACGTGAGTACAGACAGCAGTCATGTGTACAA 300
QY 301 GAAAAACAATCTCTTCGAGGTGGTGAGAAATGCGCGGTAGTGTACTACAGACAC 360
Db 301 GAAAAACAATCTCTTCGAGGTGGTGAGAAATGCGCGGTAGTGTACTACAGACAC 360
QY 361 AGGAAGCGGACTGAGAAACTTTTCAAGAAATTTCTAGCAAAATGCCAGGCACCT 420
Db 361 AGGAAGCGGACTGAGAAACTTTTCAAGAAATTTCTAGCAAAATGCCAGGCACCT 420
QY 421 ACATGCTTATGGACGTCGCTCAAGAGGACGCTGATGATTAATGATGATGATGATGAT 480
Db 421 ACATGCTTATGGACGTCGCTCAAGAGGACGCTGATGATTAATGATGATGATGATGAT 480
QY 481 TTGTGGTACAGTCAGGAGTCCCTATATGACGAAGATACAAATGCTGTATGATGAAG 540
Db 481 TTGTGGTACAGTCAGGAGTCCCTATATGACGAAGATACAAATGCTGTATGATGAAG 540
QY 541 GTAAGGTGATCCATCTGAGCAATGAGGATGATATCAAGGACTGTGGGCTCTTAGACC 600
Db 541 GTAAGGTGATCCATCTGAGCAATGAGGATGATATCAAGGACTGTGGGCTCTTAGACC 600
QY 601 AAGATGTTGACTCGATTAATGCTGAGTCAAAACGAGTGTATCTACAGACACAGTAG 660
Db 601 AAGATGTTGACTCGATTAATGCTGAGTCAAAACGAGTGTATCTACAGACACAGTAG 660
QY 661 AGGAGCTGAAGACACACCGTCAGAAACTGGAGAAATTTCTGTTAGATCTATCAGACTA 720
Db 661 AGGAGCTGAAGACACACCGTCAGAAACTGGAGAAATTTCTGTTAGATCTATCAGACTA 720
QY 721 CTTTATACAGGACCAACTGCTGCTTACACAGACACCTATACAGTGTCTCTCATCAAT 780
Db 721 CTTTATACAGGACCAACTGCTGCTTACACAGACACCTATACAGTGTCTCTCATCAAT 780
QY 781 GTGTAATAACAGAAATATCGATAGATATTTGAAATATAAATGTTAATAAACACTGGTTGA 840
Db 781 GTGTAATAACAGAAATATCGATAGATATTTGAAATATAAATGTTAATAAACACTGGTTGA 840
QY 841 AATATGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 870
Db 841 AATATGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 870
```

```
RESULT 2
PDA131051      850 bp      mRNA      linear      INV 28-MAR-2000
LOCUS          Pholas dactylus mRNA for pholasin.
DEFINITION    AJ131051
ACCESSION     AJ131051
VERSION       AJ131051.1 GI:4454298
KEYWORDS      phol gene; pholasin; photoprotein.
SOURCE        Pholas dactylus
ORGANISM      Pholas dactylus
              Eukaryota; Metazoa; Mollusca; Bivalvia; Heteroconchia; Veneroidea;
              Pholadoidea; Pholadidae; Pholas.
REFERENCE     1
AUTHORS       Dunstan,S.L., Sala-newby,G.B., Fajardo,A., Taylor,K. and
              Campbell,A.K.
TITLE         Cloning and expression of the bioluminescent photoprotein pholasin
              from the bivalve mollusc pholas dactylus
JOURNAL       J. Biol. Chem. 275, 9550-9556 (2000)
REFERENCE     2 (bases 1 to 850)
AUTHORS       Campbell,A.K.
TITLE         Direct Submission
JOURNAL       Submitted (25-NOV-1998) Campbell A.K., Medical Biochemistry,
              University of Wales College of Medicine, Heath Park, Cardiff, CF4
              4XN, UK
FEATURES      Location/Qualifiers
              source
                1. .850
                  /organism="Pholas dactylus"
                  /mol_type="mRNA"
                  /db_xref="taxon:52916"
                  /tissue_type="photophore"
                  /dev_stage="adult"
                1. .850
                  /genes="phol"
                  16..693
                  /genes="phol"
                  /function="bioluminescence photoprotein"
                  /codon_start=1
                  /evidence=experimental
                  /product="pholasin"
                  /protein_id="CAA10291.1"
                  /db_xref="GI:4454298"
                  /db_xref="SPTREMBL:O97465"
                  /translation="NACIVFVALVCLMPSGEEVOCANNWQANEYFNVDMTI
                  FYDYGAPQLYEDRALGLRIRAGFTTKRAVWNSNDQSCVTRKTIFFEVSGEI
                  ARVYRPOEDGTEKTRFKFSKMPCTYALMDVCAADDDKICETIVTVRVSLY
                  DENNGVMDGKVIPTSETIEDIKDGLLDQVDLDYTWQNECDLPDVTDEAEDTPS
                  ETCEFTW"
                16..76
                  /genes="phol"
                  /gene="phol"
                  77..690
                  /product="pholasin"
                  832..850
                  /genes="phol"
BASE COUNT    284 a   158 c   206 g   202 t
ORIGIN
Query Match    97.7%; Score 850; DB 3; Length 850;
Best Local Similarity 100.0%; Pred. No. 1.9e-203;
Matches 850; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 TCGAAAGAACAAAATGCGTTCATCGTTTCGTTGCTTCCTTCGCTCTATGCTTAATG 74
Db 1 TCGAAAGAACAAAATGCGTTCATCGTTTCGTTGCTTCCTTCGCTCTATGCTTAATG 60
QY 75 CAACCGGTTCCGGTGAGGAAGTCAATGCGCATCAATGGACACAGCTAATGAATAT 134
Db 61 CAACCGGTTCCGGTGAGGAAGTCAATGCGCATCAATGGACACAGCTAATGAATAT 120
QY 135 GTGTTCAACGTGACTGGATGACCACTTTTCATCTACGACTATGCGGCTCAAGAGCACTG 194
Db 121 GTGTTCAACGTGACTGGATGACCACTTTTCATCTACGACTATGCGGCTCAAGAGCACTG 180
```

```
QY 195 TACGAAGATCGGGCTTTGGGGCTGTGTCGGATTGAACGGCGCGCCGAGGTACACAAAA 254
D 181 TACGAAGATCGGGCTTTGGGGCTGTGTCGGATTGAACGGCGCGCCGAGGTACACAAAA 240
QY 255 GCGCTCTGGATTAACTGGAGTAACGACGACGAGTCATGTTGTAACAGAAACAAATCTTC 314
D 241 GCGCTCTGGATTAACTGGAGTAACGACGACGAGTCATGTTGTAACAGAAACAAATCTTC 300
QY 315 TTCGAGGTGTGGTGGAGAAATGCCCGGCTAGTTGACTTACAGACCAACAGGAAGCAACT 374
D 301 TTCGAGGTGTGGTGGAGAAATGCCCGGCTAGTTGACTTACAGACCAACAGGAAGCAACT 360
QY 375 GAGAAACTTTTACAGAAAAATCTCTAGCAAAATGCCGACCTTACATGCTTATGGAC 434
D 361 GAGAAACTTTTACAGAAAAATCTCTAGCAAAATGCCGACCTTACATGCTTATGGAC 420
QY 435 GTGTGCGCTACAGGACGCTGATGATAAATGCTATGCAAGGCAACATTTGTGTGACAGTC 494
D 421 GTGTGCGCTACAGGACGCTGATGATAAATGCTATGCAAGGCAACATTTGTGTGACAGTC 480
QY 495 AGGGTGTCCCTATATACGAGAAATAACAATGGTGTATGATGAAGGTAAGTGATTTCCA 554
D 481 AGGGTGTCCCTATATACGAGAAATAACAATGGTGTATGATGAAGGTAAGTGATTTCCA 540
QY 555 TCTGAGACAATCGAGGATGATATCAAGGACTGTGGCTCTTAGACCAAGATTTGAACTC 614
D 541 TCTGAGACAATCGAGGATGATATCAAGGACTGTGGCTCTTAGACCAAGATTTGAACTC 600
QY 615 GATTATAGTGGACTCAAAACAGAGTGTGATCTACACACACAGTAGAGGCTGAAGAC 674
D 601 GATTATAGTGGACTCAAAACAGAGTGTGATCTACACACACAGTAGAGGCTGAAGAC 660
QY 675 ACACGCTCAGAACTGGAGAAATCTTCTGGTAGATCTATCAGACACTTATACAGCAGG 734
D 661 ACACGCTCAGAACTGGAGAAATCTTCTGGTAGATCTATCAGACACTTATACAGCAGG 720
QY 735 ACAACTGTGCTTACAGACACACTATAACGCTGCTCATCAATAATGTTGTAACAGAAA 794
D 721 ACAACTGTGCTTACAGACACACTATAACGCTGCTCATCAATAATGTTGTAACAGAAA 780
QY 795 TAATCGATAGAAATTTGAAATAAATGTTAATAAACAACGCTGTTGAAATATGAAAAAAA 854
D 781 TAATCGATAGAAATTTGAAATAAATGTTAATAAACAACGCTGTTGAAATATGAAAAAAA 840
QY 855 AAAAAA 864
D 841 AAAAAA 850

RESULT 3
PD131053
LOCUS Pholas dactylus mRNA for pholasin, clone 5.
DEFINITION Pholas dactylus mRNA for pholasin, clone 5.
ACCESSION AJ131053
VERSION AJ131053.1 GI:4454302
KEYWORDS phol gene; pholasin; photoprotein.
SOURCE Pholas dactylus
ORGANISM Pholas dactylus
Eukaryota; Metazoa; Mollusca; Bivalvia; Heterococonchia; Veneroidea;
Pholadoidea; Pholadidae; Pholas.
REFERENCE 1
AUTHORS Dunstan,S.L., Sala-newby,G.B., Fajardo,A., Taylor,K. and
Campbell,A.K.
TITLE Cloning and expression of the bioluminescent photoprotein pholasin
from the bivalve mollusc pholas dactylus
JOURNAL J. Biol. Chem. 275, 9550-9556 (2000)
REFERENCE 2 (bases 1 to 846)
AUTHORS Campbell,A.K.
TITLE Direct Submission
JOURNAL Submitted (25-NOV-1998) Campbell A.K., Medical Biochemistry,
University of Wales College of Medicine, Heath Park, Cardiff, CF4
4XN, UK
FEATURES
```

```
source 1. .846
/organism="Pholas dactylus"
/mol_type="mRNA"
/db_xref="taxon:52916"
/clone="5"
/tissue_type="photophore"
/dev_stage="adult"
gene 1. .846
/genes="phol"
/genes="phol"
/genes="phol"
/genes="phol"
/function="bioluminescence photoprotein"
/evidence=experimental
/product="pholasin"
/protein_id="CAA10293.1"
/db_xref="GI:4454303"
/db_xref="SPTREMBL:O97465"
/translation="MACIVFVALVLCMLQPGSGEEVQCAMNTQANEYFVNDWMTI
FYDXGAEQIEDRALGICRIERAGPGTTKAVLNWNSNDPSCVTKRTITFEVGGEI
ARLDYRQDEGTEKTFKFSKMGFTMTMDVCAATDADKDCIGTITVTIVVRSLY
DEDNGVMDGKVIPTSETIEDDIKDGLLDQVVDLYTWTQNECDLPDVTDEADTPS
ETGEFW"
sig_peptide 17. .77
/genes="phol"
mat_peptide 78. .691
/genes="phol"
/genes="phol"
polya_signal 826. .846
/genes="phol"
BASE COUNT 280 a 159 c 207 g 200 t
ORIGIN
Query Match 96.0%; Score 834.8; DB 3; Length 846;
Best Local Similarity 99.2%; Pred. No. 1.3e-199;
Matches 839; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 14 CTCGGAARAACAACAATGGCTTGTATCGTTTTCGTTTCGTCGCTATGCTTAAT 73
D 1 CTCGGAARAACAACAATGGCTTGTATCGTTTTCGTTTCGTCGCTATGCTTAAT 60
QY 74 GCACCCGGGTTCCGGTGGAGAAATGCAATCGCGGATGAATGGACACAAAGCAAT 133
D 61 GCACCCGGGTTCCGGTGGAGAAATGCAATCGCGGATGAATGGACACAAAGCAAT 120
QY 134 TGTGTTCAACGTGGACTGGATGACCATTTTCATCTACGACTATGGCGCTCAAGCAACT 193
D 121 TGTGTTCAACGTGGACTGGATGACCATTTTCATCTACGACTATGGCGCTCAAGCAACT 180
QY 194 GTACGAAGATCGGGCTTTGGGCTGTGTGCGATTGAACGGCGCGCCGAGGTACCAAAA 253
D 181 GTACGAAGATCGGGCTTTGGGCTGTGTGCGATTGAACGGCGCGCCGAGGTACCAAAA 240
QY 254 AGCGCTCTGGATTAACTGGAGTAACGACGACGAGTCATGTGTAACAAGAAACAATCTT 313
D 241 AGCGCTCTGGATTAACTGGAGTAACGACGACGAGTCATGTGTAACAAGAAACAATCTT 300
QY 314 CTTGAGGTTGGTGGAGAAATTTGCCGCTAGTTGACTACAGACCAAGGAGGAGCAAC 373
D 301 CTTGAGGTTGGTGGAGAAATTTGCCGCTAGTTGACTACAGACCAAGGAGGAGCAAC 360
QY 374 TGAGAAACTTTTACAGAAAAATTTCTTAGCAAAATGCCAGGCACTTACATGCTTATGA 433
D 361 TGAGAAACTTTTACAGAAAAATTTCTTAGCAAAATGCCAGGCACTTACATGCTTATGA 420
QY 434 CGTGTGCGCTACAGGACGCTGATGATAATGCAATGCAATGCAAGGACCAATTTGTGACACT 493
D 421 CGTGTGCGCTACAGGACGCTGATGATAATGCAATGCAATGCAAGGACCAATTTGTGACACT 480
QY 494 CAGGGTGTCCCTATATGACGAGAGATAACAATGGTGTGTAATGAGTAAGTATGCTTATCC 553
D 481 CAGGGTGTCCCTATATGACGAGAGATAACAATGGTGTGTAATGAGTAAGTATGCTTATCC 540
```



Db 661 GAGAAATCTCTGCTGGTAGATCTATCAGACTACTTTTATCAGCAGGACAACCTGGTGGTTACC 720  
 QY 751 AGACACCTATAACGCTGCTCATCAATAAATGCTGTAACACAGAAAATAATCGATAGAAATATT 810  
 Db 772 AGACACCTATAACGCTGCTCATCAATAAATGCTGTAACACAGAAAATAATCGATAGAAATATT 780  
 QY 811 GAAATAATAAATGTTAATAAACAACCTGGTTGAATATGAA 848  
 Db 781 GAAATAATAAATGTTAATAAACAACCTGGTTGAATATGAA 818  
 RESULT 6  
 AX024885  
 LOCUS AX024885 816 bp DNA linear PAT 15-SEP-2000  
 DEFINITION Sequence 2 from Patent WO0028025.  
 ACCESSION AX024885  
 VERSION AX024885.1 GI:10184826  
 KEYWORDS  
 SOURCE Pholas dactylus  
 ORGANISM Pholas dactylus  
 Eukaryota; Metazoa; Mollusca; Bivalvia; Heteroconchia; Veneroidea;  
 Pholadoidea; Pholadidae; Pholas.  
 1  
 Campbell A.K.  
 TITLE Pholasin  
 JOURNAL Patent: WO 0028025-A 2 18-Mar-2000;  
 UNIV WALES MEDICINE (GB); CAMPBELL ANTHONY KEITH (GB)  
 FEATURES  
 source  
 1. 816  
 /organism="Pholas dactylus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:52916"  
 257 a 159 c 206 g 194 t  
 BASE COUNT  
 ORIGIN  
 Query Match 91.9%; Score 799.2; DB 6; Length 816;  
 Best Local Similarity 99.4%; Pred. No. 1.2e-190;  
 Matches 813; Conservative 0; Mismatches 3; Indels 2; Gaps 1  
 QY 1 GAATTCGGCAGAGTCGGGAAGAACAAAATGCTGTATCGTTTTCGTTGCTCTGTTGCG 60  
 Db 1 GAATTCGGCAGAG--GGAAAGAACAAAATGCTGTATCGTTTTCGTTGCTCTGTTGCG 58  
 QY 61 CTCATGCTTAATGCAACCGGTTCCGGTGAGGAAGTCAATGCCGATGAATTTGGACAC 120  
 Db 59 CTCATGCTTAATGCAACCGGTTCCGGTGAGGAAGTCAATGCCGATGAATTTGGACAC 118  
 QY 121 AAGCTAATGAATATGTGTTCAACGTGGAGTGGATGACCAATTTCACTACGACTATGGCG 180  
 Db 119 AAGCTAATGAATATGTGTTCAACGTGGAGTGGATGACCAATTTCACTACGACTATGGCG 178  
 QY 181 CTCAGAGCAACTGTACGAAGATCGGGCTTTGGGGCTGTGTCGATTGAACGGGCGGGCC 240  
 Db 179 CTCAGAGCAACTGTACGAGGATCGGGCTTTGGGGCTGTGTCGATTGAACGGGCGGGCC 238  
 QY 241 CAGGTACCAAAAAGCCGCTCGGATTAACGTGGATTAACGACGACGAGTCATGTGTACAA 300  
 Db 239 CAGGTACCAAAAAGCCGCTCGGATTAACGTGGATTAACGACGACGAGTCATGTGTACAA 298  
 QY 301 GAAAAACAATCTTCTTCGAGGTTGTTGGAGAAATTCGCCGGCTACTTGACTACAGACAC 360  
 Db 299 GAAAAACAATCTTCTTCGAGGTTGTTGGAGAAATTCGCCGGCTACTTGACTACAGACAC 358  
 QY 361 AGGAAGACCGAATGAGAAAATCTTTACAGAAAATCTCTAGCAAAAATGCCAGGCACATT 420  
 Db 359 AGGAAGACCGAATGAGAAAATCTTTACAGAAAATCTCTAGCAAAAATGCCAGGCACATT 418  
 QY 421 ACATGCTTATGGACGCTGCGCTACAAGGGACGCTGATGATAAATGCATCGAAGGCACAA 480  
 Db 419 ACATGCTTATGGACGCTGCGCTACAAGGGACGCTGATGATAAATGCATCGAAGGCACAA 478  
 QY 481 TTGTGGTGACAGTCAGGGGTGCCCTTATATGACGAAGATAACAATGGTGAATGGATGAAG 540















```

BASE COUNT      566 a      119 c      221 g      810 t
ORIGIN

Query Match      5.4%; Score 46.8; DB 3; Length 1716;
Best Local Similarity 58.7%; Pred. No. 0.58;
Matches 81; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 729 AGCAGGACAACTGGTCGTACAGACACCTATACCGTCTCCTCAATCAATATGTGTAAA 788
Db 150 ATCAAGTAAATAGTCATACCACTCTTACAGAAATTTAAAGAAATATTTCTCACTCAA 91
QY 789 CAGAAATATCGATAGAAATATTGAAATATAAAATGTTAATAAACACTGGTTGAAATATGAA 848
Db 90 AAAAAATACATATATTTTTATATAAAAAAATTTATAAAAAAATTTTATTACATAAA 31
QY 849 AAAAAAATAAAAAAACT 866
Db 30 AAAAAATAAAAAAACT 13

```

Search completed: August 22, 2003, 05:52:56  
Job time : 3478 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2003, 01:22:03 ; Search time 299 Seconds  
(without alignments)  
7854.552 Million cell updates/sec

Title: US-09-831-142B-1  
Perfect score: 870  
Sequence: 1 gaattcgacagtcggaa.....aaaaaataaaactogag 870

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N.Geneseq\_19Jun03:\*  
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*  
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	870	100.0	870	21	AA15211	CDNA sequence from
2	834.8	96.0	852	21	AA15213	CDNA sequence from
3	818	94.0	818	21	AA15214	PCR amplified geno
4	799.2	91.9	815	21	AA15212	CDNA sequence from
5	725	83.3	726	21	AA15215	PCR amplified geno
6	47.4	5.4	9810	24	ABL32427	Human immune syste
7	47	5.4	758	22	ABA06495	Human CDNA SEQ ID
8	47	5.4	758	24	ABV83832	Human polynucleoti

C	9	46.2	5.3	9741	24	ABL33323	Human immune syste
	10	45.6	5.2	360	23	ABV59054	Human prostate exp
	11	45.6	5.2	3028	24	ABA93725	Human differentiat
	12	45.4	5.2	488	23	ABV58909	Human prostate exp
	13	45.4	5.2	2091	24	AA150827	Human cancer statu
	14	45.4	5.2	2710	21	AA54132	Breast cancer prot
	15	45	5.2	17294	24	ABL32987	Human immune syste
C	16	44.8	5.1	6211	24	ABL32806	Human immune syste
C	17	44.2	5.1	407	22	AA183508	Human polynucleoti
C	18	44	5.1	6040	24	ABL32067	Human immune syste
C	19	43.8	5.0	363	25	ABX46522	Bovine EST associa
C	20	43.8	5.0	929	24	ABQ54553	Human ovarian anti
C	21	43.4	5.0	11745	24	ABK28332	DNA transcription
C	22	43.2	5.0	1315	22	AAF81788	Human secreted pro
C	23	43.2	5.0	1315	25	ACC50854	Human secreted pro
C	24	43.2	5.0	1317	25	ACC50674	Human colon cancer
C	25	43.2	5.0	1321	21	AAC98225	Human secreted pro
C	26	43.2	5.0	2185	24	AA145822	Human secreted pro
C	27	43.2	5.0	2852	24	AEQ73250	Human MAP kinase p
C	28	43.2	5.0	5649	22	AA546384	Tumour suppressor
C	29	43.2	5.0	5649	24	ABK40008	Human chemically p
C	30	43.2	5.0	5649	24	ABL32849	Human immune syste
C	31	43	4.9	1350	24	ABL90008	Human polynucleoti
C	32	43	4.9	1904	22	AAS21293	Human CDNA sequenc
C	33	43	4.9	1904	25	ACA03652	Human CDNA encodin
C	34	43	4.9	1904	25	ACA04073	Human CDNA encodin
C	35	43	4.9	1904	25	ABX89190	DNA encoding novel
C	36	43	4.9	2665	19	AAV32592	Schwannomin-bindin
C	37	43	4.9	6236	24	ABN80203	Human chemically m
C	38	43	4.9	6236	24	ABL33631	Human immune syste
C	39	42.8	4.9	203	24	AD31846	Human pancreatic t
C	40	42.8	4.9	276	25	ABX44366	Bovine EST associa
C	41	42.8	4.9	383	22	AA188525	Human polynucleoti
C	42	42.8	4.9	524	25	ABF22885	Breast cancer mark
C	43	42.8	4.9	826	22	AAF72823	Secreted protein g
C	44	42.8	4.9	935	21	AAC95101	Cat flea hindgut a
C	45	42.8	4.9	1635	25	ABX12483	CDNA encoding rice

ALIGNMENTS

RESULT 1  
AA15211  
ID AA15211 standard; cDNA; 870 BP.

XX	AC	AA15211;				
XX	DF	04-SEP-2000 (first entry)				
XX	DE	CDNA sequence from clone 40 encoding apopholasin.				
XX	KW	Bivalve mollusc; apopholasin; bioluminescent oxidative indicator protein;				
XX	KW	BOIP; light emission; pholasin; oxygen; chemiluminesce; cancer cell;				
XX	KW	hyperactive cell; rheumatoid arthritis; inflammatory disease; ss.				
XX	OS	Pholas dactylus.				
XX	XX	Key	Location/Qualifiers			
XX	FT	CDS	30..707			
XX	FT		/*tag= a			
XX	FT		/product= "apopholasin"			
XX	PN	W0200028025-A1.				
XX	PD	18-MAY-2000.				
XX	PF	05-NOV-1999;	99WO-GB03654.			
XX	PR	07-NOV-1998;	98GB-0024357.			
XX	PA	(UJWA-) UNIV WALES COLLEGE OF MEDICINE.				

PI Campbell AK;  
 XX WPI: 2000-387420/33.  
 DR P-PSDB; AAY93251.  
 XX  
 PT Novel recombinant nucleic acid molecules that encode the apophoprotein  
 of pholasin or its homologous sequence useful for detecting location  
 PT and measurement of oxygen and its metabolites in living cells and  
 PT organs  
 XX  
 PS Claim 3; Fig 1; 49pp; English.  
 XX  
 CC The present sequence encodes a bivalve mollusc apopholasin protein.  
 CC The protein is a bioluminescent oxidative indicator protein (BOIP).  
 CC Changes in light emission of pholasin enable oxygen or its metabolites  
 CC to be detected and quantified in live cells, organellies or on the outer  
 CC or inner surface of the plasma membrane, or within an organ of a live  
 CC organism without the need to break them open or the need to separate  
 CC bound and free fractions. This also enables an enzyme producing oxygen  
 CC or one of its metabolites to be detected and quantified. The BOIP  
 CC is used for the detection, diagnosis or measurement of oxygen or its  
 CC metabolites intracellularly or extracellularly. The BOIP includes a  
 CC signal peptide whose target is set to a predetermined extra or  
 CC intracellular site. The light emission preferably takes place in the  
 CC absence of the luciferase. Pholasin is also useful as a protein or a DNA  
 CC label or in genetic entertainment which involves adding pholasin to drink  
 CC such as beer, cola, soft drinks and spirits to make them glow since  
 CC pholasin is able to chemiluminesce at a wide range of pH (3-10). It can  
 CC also be added to foodstuffs and in a wide range of toys and other  
 CC entertaining devices. BOIP nucleic acids can be used for detection and  
 CC location of abnormal cells such as cancer cells, hyperactive cells in  
 CC rheumatoid arthritis and other inflammatory diseases, cells infected with  
 CC a pathogen, damaged cells, and measurement and location of enzymes.  
 XX  
 SQ Sequence 870 BP; 289 A; 163 C; 213 G; 205 T; 0 other;  
 Query Match 100.0%; Score 870; DB 21; Length 870;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-206;  
 Matches 870; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GAATTCGCGACGAGTCGAAAGAAACAAATGCGTTGATCGTTTTCGTTCTGCTGCG 60  
 DB 1 GAATTCGCGACGAGTCGAAAGAAACAAATGCGTTGATCGTTTTCGTTCTGCTGCG 60  
 QY 61 CTCATGCTTAATGCAACCGGTTCCGCTGAGGAGTACAAATGCGGATGAATGCGAC 120  
 DB 61 CTCATGCTTAATGCAACCGGTTCCGCTGAGGAGTACAAATGCGGATGAATGCGAC 120  
 QY 121 AAGCTAATGATATGTTCAACGCTGACTGATGACCAATTTTCATCTACGACTATGGCG 180  
 DB 121 AAGCTAATGATATGTTCAACGCTGACTGATGACCAATTTTCATCTACGACTATGGCG 180  
 QY 181 CTCAGAGCAACTCTACGAGATCGGGCTTTGGGGCTGTGTCGGATTGAACGGCCGGCC 240  
 DB 181 CTCAGAGCAACTCTACGAGATCGGGCTTTGGGGCTGTGTCGGATTGAACGGCCGGCC 240  
 QY 241 CAGGTACACAAAACCGCTCGATTACTGAGTACACGACGACGACGATCTGTACAA 300  
 DB 241 CAGGTACACAAAACCGCTCGATTACTGAGTACACGACGACGACGATCTGTACAA 300  
 QY 301 GAAAAACAATCTCTTCGAGTTGTTGGAGAAATTCGCCGCTAGTGTGACTACAGACCAC 360  
 DB 301 GAAAAACAATCTCTTCGAGTTGTTGGAGAAATTCGCCGCTAGTGTGACTACAGACCAC 360  
 QY 361 AGAAGACGGAAGTACGAAACCTTTTACAGAAATTTCTTAGCAAAATGCCAGGCACTT 420  
 DB 361 AGAAGACGGAAGTACGAAACCTTTTACAGAAATTTCTTAGCAAAATGCCAGGCACTT 420  
 QY 421 ACATGCTTATGAGTGTGGCTGACAGGAGCGCTGATGATAAATGCATCGAAGGCACAA 480  
 DB 421 ACATGCTTATGAGTGTGGCTGACAGGAGCGCTGATGATAAATGCATCGAAGGCACAA 480  
 QY 481 TTGTGTCGACAGTCAGGCTGCTCCCTATATGACGAAGATAACAATGCTGTATGATGAAG 540

DB 481 TTGTGTCGACAGTCAGGCTGCTCCCTATATGACGAAGATAACAATGCTGTATGATGAAG 540  
 QY 541 GTAAGTGATTCATCTGAGCAATCGAGGATGATATCAAGACTGTGGCTCTTAGACC 600  
 DB 541 GTAAGTGATTCATCTGAGCAATCGAGGATGATATCAAGACTGTGGCTCTTAGACC 600  
 QY 601 AAGATGTTGAACCTCGATTATACGTGGACTCAAAACGAGTGTGATCTACACACACAGTAG 660  
 DB 601 AAGATGTTGAACCTCGATTATACGTGGACTCAAAACGAGTGTGATCTACACACACAGTAG 660  
 QY 661 ACGAGCTGAGACACACACCTGAGAACTGAGAAATCTCTGCTGATCTATCAGACTA 720  
 DB 661 ACGAGCTGAGACACACACCTGAGAACTGAGAAATCTCTGCTGATCTATCAGACTA 720  
 QY 721 CTTTATCAGCAGACAACTGCTGCTTACCAGACACCTATACGCTGCTCTCATCAATAAT 780  
 DB 721 CTTTATCAGCAGACAACTGCTGCTTACCAGACACCTATACGCTGCTCTCATCAATAAT 780  
 QY 781 GTGTAACACAGAAATATCGATAGATAATTTGAAATATAATGTTAATAACACTGTTGA 840  
 DB 781 GTGTAACACAGAAATATCGATAGATAATTTGAAATATAATGTTAATAACACTGTTGA 840  
 QY 841 AATATGAAAAAATAAAAAAATACTCGAG 870  
 DB 841 AATATGAAAAAATAAAAAAATACTCGAG 870  
 RESULT 2  
 AAL5213  
 ID AAL5213 standard; cDNA; 852 BP.  
 XX  
 AC AAL5213;  
 XX  
 DT 04-SEP-2000 (first entry)  
 XX  
 DE cDNA sequence from clone 5 encoding apopholasin.  
 XX  
 KW Bivalve mollusc; apopholasin; bioluminescent oxidative indicator protein;  
 KW BOIP; light emission; pholasin; oxygen; chemiluminescence; cancer cell;  
 KW hyperactive cell; rheumatoid arthritis; inflammatory disease; ss.  
 XX  
 OS Pholas dactylus.  
 XX  
 FH Key  
 FT CDS 17..694  
 FT /tag= a  
 FT /product= "apopholasin"  
 XX  
 XX WO200028025-A1.  
 XX 18-MAY-2000.  
 XX  
 XX 05-NOV-1999; 99WO-GB03654.  
 XX  
 XX 07-NOV-1998; 98GB-0024357.  
 XX  
 XX (UIWA-) UNIV WALES COLLEGE OF MEDICINE.  
 XX  
 XX Campbell AK;  
 XX  
 XX WPI: 2000-387420/33.  
 XX P-PSDB; AAY93251.  
 XX  
 PT Novel recombinant nucleic acid molecules that encode the apophoprotein  
 of pholasin or its homologous sequence useful for detecting location  
 PT and measurement of oxygen and its metabolites in living cells and  
 PT organs  
 XX  
 PS Claim 3; Fig 1; 49pp; English.  
 XX  
 CC The present sequence encodes a bivalve mollusc apopholasin protein.  
 CC The protein is a bioluminescent oxidative indicator protein (BOIP).

CC changes in light emission of pholasin enable oxygen or its metabolites  
CC to be detected and quantified in live cells, organelles or on the outer  
CC or inner surface of the plasma membrane, or within an organ of a live  
CC organism without the need to break them open or the need to separate  
CC bound and free fractions. This also enables an enzyme producing oxygen  
CC or one of its metabolites to be detected and quantified. The BOIP  
CC is used for the detection, diagnosis or measurement of oxygen or its  
CC metabolites intracellularly or extracellularly. The BOIP includes a  
CC signal peptide whose target is set to a predetermined extra or  
CC intracellular site. The light emission preferably takes place in the  
CC absence of the luciferase. Pholasin is also useful as a protein or a DNA  
CC label or in genetic entertainment which involves adding pholasin to drink  
CC such as beer, cola, soft drinks and spirits to make them glow since  
CC pholasin is able to chemiluminesce at a wide range of pH (3-10). It can  
CC also be added to foodstuffs and in a wide range of toys and other  
CC entertaining devices. BOIP nucleic acids can be used for detection and  
CC location of abnormal cells such as cancer cells, hyperactive cells in  
CC rheumatoid arthritis and other inflammatory diseases, cells infected with  
CC a pathogen, damaged cells, and measurement and location of enzymes.

XX Sequence 852 BP; 281 A; 161 C; 209 G; 201 T; 0 other;

Query Match 96.0%; Score 834.8; DB 21; Length 852;  
Best Local Similarity 99.2%; Pred. No. 6,4e-198;  
Matches 839; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 14 GTGGAAGAAGCAAAATGGCTGTATCGTTTTCGTTGCTTCCTGCTATGCTTAAT 73  
DB 1 GTGGAAGAAGCAAAATGGCTGTATCGTTTTCGTTGCTTCCTGCTATGCTTAAT 60  
QY 74 GAACCGGTTCGCTGAGGAGTCAATGCGCGATGAATGGACACAGCTATGATA 133  
DB 61 GAACCGGTTCGCTGAGGAGTCAATGCGCGATGAATGGACACAGCTATGATA 120  
QY 134 TGTGTTCAACGTGGAGTGGATGACCAATTTTCATCTACGACTATGCGCTCAAGACCACT 193  
DB 121 TGTGTTCAACGTGGAGTGGATGACCAATTTTCATCTACGACTATGCGCTCAAGACCACT 180  
QY 194 GTACGAGATCGGCTTTGGGCTGTGCTGATGAAAGCGGCGCCAGGTTACCAAAA 253  
DB 181 GTACGAGATCGGCTTTGGGCTGTGCTGATGAAAGCGGCGCCAGGTTACCAAAA 240  
QY 254 AGCGCTCTGGATTAACTGGAGTAACGACAGCAGTATGTTAAACAGAAAAAACAATCTT 313  
DB 241 AGCGCTCTGGATTAACTGGAGTAACGACAGCAGTATGTTAAACAGAAAAAACAATCTT 300  
QY 314 CTTGCGGTTGTTGGAGAAATTCGCGGCTAGTTCAGTACAGCCACAGAGAGCGGAC 373  
DB 301 CTTGCGGTTGTTGGAGAAATTCGCGGCTAGTTCAGTACAGCCACAGAGAGCGGAC 360  
QY 374 TGAGAAAACCTTTTACAAGAAAATTCCTAGCAAAATGCCAGGCACTTATGCTTATGGA 433  
DB 361 TGAGAAAACCTTTTACAAGAAAATTCCTAGCAAAATGCCAGGCACTTATGCTTATGGA 420  
QY 434 CGTGTGCGCTACAGGAGCAGCTGATGATTAATGCAATGCAAGGACCAATTTGCTGGACAGT 493  
DB 421 CGTGTGCGCTACAGGAGCAGCTGATGATTAATGCAATGCAAGGACCAATTTGCTGGACAGT 480  
QY 494 CAGGCTGTCCCTATATGACGAAGATTAACAAATGGTGAATGAAGTAAAGTGATGCC 553  
DB 481 CAGGCTGTCCCTATATGACGAAGATTAACAAATGGTGAATGAAGTAAAGTGATGCC 540  
QY 554 ATCTGAGCAATPCGAGGATGATATCAAGGACGTGGGCTCTTAGACCAAGATGTTGAAGT 613  
DB 541 ATCTGAGCAATPCGAGGATGATATCAAGGACGTGGGCTCTTAGACCAAGATGTTGAAGT 600  
QY 614 CGATTTATGCTGAGTCAAAACGAGTGTGATCTACACAGACAGTACAGAGCGCTGAAGA 673  
DB 601 CGATTTATGCTGAGTCAAAACGAGTGTGATCTACACAGACAGTACAGAGCGCTGAAGA 660  
QY 674 CACACCGTCAAGAACTGGAATTTCTGTTGATGATCTATCAGACTACTTTTATCAGCAG 733  
DB 661 CACACCGTCAAGAACTGGAATTTCTGTTGATGATCTATCAGAGCCACTTTATCAGCAG 720

QY 734 GACAACTGGTGGTTCACACACCTATAAGCTGCTCCTCATCAATAATGCTGTAAACAGAA 793  
DB 721 GACAACTGGTGGTTCACACACCTATAAGCTGCTCCTCATCAATAATGCTGTAAACAGAA 780  
QY 794 ATAATCGATAGATATTTGAATAAATAAATGTTAATAAACACTGGTTGAAATATGAAAAAA 853  
DB 781 ATAATCGATAGATATTTGAATAAATAAATGTTAATAAACACTGGTTGAAATATGAAAAAA 840  
QY 854 AAAAAA 859  
DB 841 AAAAAA 846  
RESULT 3  
AAAL5214  
ID AAAL5214 standard; CDNA; 818 BP.  
XX AAAL5214;  
XX 04-SEP-2000 (first entry)  
XX PCR amplified genomic sequence of apopholasin DNA.  
DE Bivalve mollusc; apopholasin; bioluminescent oxidative indicator protein;  
XX BOIP; light emission; pholasin; oxygen; chemiluminesce; cancer cell;  
XX hyperactive cell; rheumatoid arthritis; inflammatory disease; ss.  
OS Pholasin dactylus.  
XX WO200028025-A1.  
XX 18-MAY-2000.  
XX 05-NOV-1999; 99WO-GB03654.  
XX 07-NOV-1998; 98GB-0024357.  
XX (UYWA-) UNIV WALES COLLEGE OF MEDICINE.  
XX Campbell AK;  
XX WPI; 2000-387420/33.  
XX Novel recombinant nucleic acid molecules that encode the apophoprotein  
XX of pholasin or its homologous sequence useful for detecting location  
XX and measurement of oxygen and its metabolites in living cells and  
XX organs -  
XX Disclosure: Fig 6; 49pp; English.

XX The present sequence represents a PCR amplified genomic sequence of  
XX apopholasin DNA. The protein is a bioluminescent oxidative indicator  
XX protein (BOIP). Changes in light emission of pholasin enable oxygen or  
XX its metabolites to be detected and quantified in live cells, organelles  
XX or on the outer or inner surface of the plasma membrane, or within an  
XX organ of a live organism without the need to break them open or the need  
XX to separate bound and free fractions. This also enables an enzyme  
XX producing oxygen or one of its metabolites to be detected and quantified.  
XX The BOIP is used for the detection, diagnosis or measurement of oxygen or  
XX its metabolites intracellularly or extracellularly. The BOIP includes a  
XX signal peptide whose target is set to a predetermined extra or  
XX intracellular site. The light emission preferably takes place in the  
XX absence of the luciferase. Pholasin is also useful as a protein or a DNA  
XX label or in genetic entertainment which involves adding pholasin to drink  
XX such as beer, cola, soft drinks and spirits to make them glow since  
XX pholasin is able to chemiluminesce at a wide range of pH (3-10). It can  
XX also be added to foodstuffs and in a wide range of toys and other  
XX entertaining devices. BOIP nucleic acids can be used for detection and  
XX location of abnormal cells such as cancer cells, hyperactive cells in  
XX rheumatoid arthritis and other inflammatory diseases, cells infected with  
XX a pathogen, damaged cells, and measurement and location of enzymes.

## RESULT 4



181 CTCAGAGCAACTGTACAGAGATCGGGCTTTGGGCTGTGTGCGATTGAACGGCGCGCC 240  
 179 CTCAGAGCAACTGTACAGAGATCGGGCTTTGGGCTGTGTGCGATTGAACGGCGCGCC 238  
 241 CAGGTACCAAAAGCCGTCTGGATTAACCTGAGTAAGACACGCGAGTATGTGTAAACA 300  
 239 CAGGTACCAAAAGCCGTCTGGATTAACCTGAGTAAGACACGCGAGTATGTGTAAACA 298  
 301 GAAAAACAATCTCTTCAGGTTGGTGGAGAAATTCGCCGGTAGTGTACTACAGCAC 360  
 299 GAAAAACAATCTCTTCAGGTTGGTGGAGAAATTCGCCGGTAGTGTACTACAGCAC 358  
 361 AGAAGACGGAAGTACAGAAACTTTTACAGAAATTTCTTACAGAAATTTCCAGCAACT 420  
 359 AGAAGACGGAAGTACAGAAACTTTTACAGAAATTTCTTACAGAAATTTCCAGCAACT 418  
 421 ACATGCTTATGACGCTGTGCGCTACAGGAGCGCTGATGAATAATGTCATCGAGCAAA 480  
 419 ACATGCTTATGACGCTGTGCGCTACAGGAGCGCTGATGAATAATGTCATCGAGCAAA 478  
 481 TTGTGTGACAGTCAAGGCTGTCCCTATATGACAGAAATTAACAATGCTGTAAATGGATGA 540  
 479 TTGTGTGACAGTCAAGGCTGTCCCTATATGACAGAAATTAACAATGCTGTAAATGGATGA 538  
 541 GTPAGGTGATTCATCTGACAGCAATCGAGATGATATCAAGGCTGTGGGCTCTTAGACC 600  
 539 GTPAGGTGATTCATCTGACAGCAATCGAGATGATATCAAGGCTGTGGGCTCTTAGACC 598  
 601 AAGATCTTGAAGTATGATGAGGACTCAAGAGGCTGATGATCAAGGCTGTGGGCTCTTAGACC 660  
 599 AAGATCTTGAAGTATGATGAGGACTCAAGAGGCTGATGATCAAGGCTGTGGGCTCTTAGACC 658  
 661 ACAGGCTGAAGACACACCGCTCAAGAACTGGAGAAATTTCTTGTAGATCTATCAAGACTA 720  
 659 ACAGGCTGAAGACACACCGCTCAAGAACTGGAGAAATTTCTTGTAGATCTATCAAGACTA 718  
 721 CTTTATACAGCAGGCAACTGGTGTGTTACAGACACCTATACCTGCTCAATCAATAT 780  
 719 CTTTATACAGCAGGCAACTGGTGTGTTACAGACACCTATACCTGCTCAATCAATAT 778  
 781 GTGTAAACAGAAATTAATGATAGATATTTGAAATAA 818  
 779 GTGTAAACAGAAATTAATGATAGATATTTGAAATAA 816

RESULT 5  
 ARA15215  
 ID ARA15215 standard; cDNA; 726 BP.  
 XX  
 AC ARA15215;  
 DT  
 XX 04-SEP-2000 (first entry)  
 DE PCR amplified genomic sequence of apopholasin DNA.  
 XX  
 KW Bivalve mollusc; apopholasin; bioluminescent oxidative indicator protein;  
 KW BOIP; light emission; pholasin; oxygen; chemiluminesce; cancer cell;  
 KW hyperactive cell; rheumatoid arthritis; inflammatory disease; ss.  
 XX  
 OS Pholas dactylus.  
 XX  
 FN WO200028025-A1.  
 XX  
 PD 18-MAY-2000.  
 XX  
 PF 05-NOV-1999; 99WO-GB03654.  
 XX  
 PR 07-NOV-1998; 98GB-0024357.  
 XX  
 PA (UYWA-) UNIV WALES COLLEGE OF MEDICINE.  
 XX  
 PI Campbell AK;  
 XX

DR WPI; 2000-387420/33.  
 XX Novel recombinant nucleic acid molecules that encode the apophoprotein  
 PT of pholasin or its homologous sequence useful for detecting location  
 PT and measurement of oxygen and its metabolites in living cells and  
 XX organs  
 PS Disclosure; Fig 6; 49pp; English.  
 XX The present sequence represents a PCR amplified genomic sequence of  
 CC apopholasin DNA. The protein is a bioluminescent oxidative indicator  
 CC protein (BOIP). Changes in light emission of pholasin enable oxygen or  
 CC its metabolites to be detected and quantified in live cells, organelles  
 CC on the outer or inner surface of the plasma membrane, or within an  
 CC organ of a live organism without the need to break them open or the need  
 CC to separate bound and free fractions. This also enables an enzyme  
 CC producing oxygen or one of its metabolites to be detected and quantified.  
 CC The BOIP is used for the detection, diagnosis or measurement of oxygen or  
 CC its metabolites intracellularly or extracellularly. The BOIP includes a  
 CC signal peptide whose target is set to a predetermined extra or  
 CC intracellular site. The light emission preferably takes place in the  
 CC absence of the luciferase. Pholasin is also useful as a protein or a DNA  
 CC label or in genetic entainment which involves adding pholasin to drink  
 CC such as beer, cola, soft drinks and spirits to make them glow since  
 CC pholasin is able to chemiluminesce at a wide range of pH (3-10). It can  
 CC also be added to foodstuffs and in a wide range of toys and other  
 CC entertaining devices. BOIP nucleic acids can be used for detection and  
 CC location of abnormal cells such as cancer cells, hyperactive cells in  
 CC rheumatoid arthritis and other inflammatory diseases, cells infected with  
 CC a pathogen, damaged cells, and measurement and location of enzymes.  
 XX Sequence 726 BP; 226 A; 145 C; 185 G; 169 T; 1 other;  
 SQ  
 Query Match 83.3%; Score 725; DB 21; Length 726;  
 Best Local Similarity 99.9%; Pred. No. 1.3e-170;  
 Matches 725; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 64 TATGCTTAAATCAACCGGTTCCGTTGGAGTAACATCGCGATGATGGACACAG 123  
 DB 1 TATGCTTAAATCAACCGGTTCCGTTGGAGTAACATCGCGATGATGGACACAG 60  
 QY 124 CTAATGAATATGTCTCAACGTGGATGACCATTTTCATCTACGACTATGCGCTC 183  
 DB 61 CTAATGAATATGTCTCAACGTGGATGACCATTTTCATCTACGACTATGCGCTC 120  
 QY 184 AAGACCAACTGTACAGATCGGGCTTTGGGCTGTGTCTCGATGACGGCGCCCGCAG 243  
 DB 121 AAGACCAACTGTACAGATCGGGCTTTGGGCTGTGTCTCGATGACGGCGCCCGCAG 180  
 QY 244 GTACCAACAAAACCGCTCTGGATTAACTGAGTAACAGACACGACGATCATGTGTAAACAGAA 303  
 DB 181 GTACCAACAAAACCGCTCTGGATTAACTGAGTAACAGACACGACGATCATGTGTAAACAGAA 240  
 QY 304 AACCAATCTTCTCGAGTTGGTGGAGAAATTCGCCGGTAGTGTGACTACAGACACAGG 363  
 DB 241 AACCAATCTTCTCGAGTTGGTGGAGAAATTCGCCGGTAGTGTGACTACAGACACAGG 300  
 QY 364 AAGACGGAAGTACAGAAACTTTTACAGAAATTTCTCTAGCAAAATGCCAGGACATTACA 423  
 DB 301 AAGACGGAAGTACAGAAACTTTTACAGAAATTTCTCTAGCAAAATGCCAGGACATTACA 360  
 QY 424 TGTATTGACGCTGTGGCTTACAGGAGCGCTGATGATAAATCATGTCGAGGACCAATG 483  
 DB 361 TGTATTGACGCTGTGGCTTACAGGAGCGCTGATGATAAATCATGTCGAGGACCAATG 420  
 QY 484 TGGTGACAGTCAAGGCTGTCCCTATATGACGAGATACAAATGCTGTAAATGATGAGGTA 543  
 DB 421 TGGTGACAGTCAAGGCTGTCCCTATATGACGAGATACAAATGCTGTAAATGATGAGGTA 480  
 QY 544 AGGTGATTCCTATCTGAGACAAATCGAGGATGATATCAAGGACTGTGGCTCTTAGACCAAG 603  
 DB 481 AGGTGATTCCTATCTGAGACAAATCGAGGATGATATCAAGGACTGTGGCTCTTAGACCAAG 540

[illegible]

PR	08-NOV-2000;	2000US-246610P.
PR	08-NOV-2000;	2000US-246611P.
PR	08-NOV-2000;	2000US-246613P.
PR	17-NOV-2000;	2000US-249207P.
PR	17-NOV-2000;	2000US-249208P.
PR	17-NOV-2000;	2000US-249309P.
PR	17-NOV-2000;	2000US-249310P.
PR	17-NOV-2000;	2000US-249311P.
PR	17-NOV-2000;	2000US-249312P.
PR	17-NOV-2000;	2000US-249313P.
PR	17-NOV-2000;	2000US-249314P.
PR	17-NOV-2000;	2000US-249315P.
PR	17-NOV-2000;	2000US-249316P.
PR	17-NOV-2000;	2000US-249317P.
PR	17-NOV-2000;	2000US-249318P.
PR	17-NOV-2000;	2000US-249344P.
PR	17-NOV-2000;	2000US-249345P.
PR	17-NOV-2000;	2000US-249364P.
PR	17-NOV-2000;	2000US-249365P.
PR	17-NOV-2000;	2000US-249397P.
PR	17-NOV-2000;	2000US-249399P.
PR	17-NOV-2000;	2000US-249300P.
PR	01-DEC-2000;	2000US-250160P.
PR	01-DEC-2000;	2000US-250391P.
PR	05-DEC-2000;	2000US-251030P.
PR	05-DEC-2000;	2000US-251388P.
PR	05-DEC-2000;	2000US-256719P.
PR	06-DEC-2000;	2000US-256147P.
PR	08-DEC-2000;	2000US-251856P.
PR	08-DEC-2000;	2000US-251868P.
PR	08-DEC-2000;	2000US-251869P.
PR	08-DEC-2000;	2000US-251989P.
PR	08-DEC-2000;	2000US-251990P.
PR	11-DEC-2000;	2000US-254097P.
PR	05-JAN-2001;	2001US-259678P.
XX		
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA		
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI; 2001-476161/51.	
DR	P-PSDB; ABB10273.	
XX		
PT	Isolated nucleic acid molecule encoding an inflammation-associated	
PT	polypeptide is used in preventing, treating or ameliorating a medical	
PT	condition -	
XX		
XX	Claim 1; SEQ ID NO: 161; 859pp + Sequence Listing; English.	
XX		
CC	The present invention provides human cDNAs, proteins and related genomic	
CC	DNA's. These can be used in the treatment of neural, immune system,	
CC	muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,	
CC	renal and proliferative disorders and inflammation. The present sequence	
XX	is a cDNA of the invention.	
XX		
SQ	Sequence 758 BP; 287 A; 164 C; 184 G; 123 T; 0 other;	

	Query Match	5.4%;	Score 47;	DB 22;	Length 758;
	Best Local Similarity	54.3%;	Prod. No. 0.063;		
	Matches 95;	Conservative	0;	Mismatches 80;	Indels 0;
	Gaps	0;			
QY	630	GGAGAAATCTCTCGTGTAGACTATATCAGACACTCTTTTATCAGCAGGACGACACTGGTGCCTTAC	749		
Db	564	GGAGGTGTATCAGGCAGCTCTGGCCACCTGGGTTTATACCCCAATATAGACGACACAGGAT	623		
QY	750	CAGACACTATACGCTGTCCCTCACTCAATAATGTGTAAACAGAAATATCGATAGAAATAT	809		
Db	624	ATGACACCCCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	683		
QY	810	TGAAATATAAATGTWTATTAACACTGCTTGAATATGAAAAAATAAAAAAATAAAAAA	864		
Db	684	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	738		

RESULT 8  
 ABV83832  
 ID ABV83832 standard; cDNA; 758 BP.  
 AC ABV83832;  
 XX  
 DT 09-DEC-2002 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 161.  
 XX  
 KW Human; neurotropic; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
 KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;  
 KW antineumatic; hepatotoxic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine;  
 KW gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002090672-A1.  
 XX  
 PD 11-JUL-2002.  
 XX  
 PF 17-JAN-2001; 2001US-0764853.  
 XX  
 PR 31-JAN-2000; 2000US-179065P.  
 PR 04-FEB-2000; 2000US-180628P.  
 PR 28-JUN-2000; 2000US-214886P.  
 PR 07-JUL-2000; 2000US-216647P.  
 PR 07-JUL-2000; 2000US-216880P.  
 PR 11-JUL-2000; 2000US-217487P.  
 PR 11-JUL-2000; 2000US-217496P.  
 PR 14-JUL-2000; 2000US-218230P.  
 PR 26-JUL-2000; 2000US-220933P.  
 PR 26-JUL-2000; 2000US-220964P.  
 PR 14-AUG-2000; 2000US-224518P.  
 PR 14-AUG-2000; 2000US-224519P.  
 PR 14-AUG-2000; 2000US-225267P.  
 PR 14-AUG-2000; 2000US-225268P.  
 PR 14-AUG-2000; 2000US-225270P.  
 PR 14-AUG-2000; 2000US-225447P.  
 PR 14-AUG-2000; 2000US-225757P.  
 PR 14-AUG-2000; 2000US-225758P.  
 PR 22-AUG-2000; 2000US-226868P.  
 PR 01-SEP-2000; 2000US-228924P.  
 PR 01-SEP-2000; 2000US-228927P.  
 PR 01-SEP-2000; 2000US-229343P.  
 PR 01-SEP-2000; 2000US-229344P.  
 PR 01-SEP-2000; 2000US-229345P.  
 PR 05-SEP-2000; 2000US-229509P.  
 PR 05-SEP-2000; 2000US-229513P.  
 PR 08-SEP-2000; 2000US-229513P.  
 PR 21-SEP-2000; 2000US-234223P.  
 PR 21-SEP-2000; 2000US-234274P.  
 PR 25-SEP-2000; 2000US-234997P.  
 PR 27-SEP-2000; 2000US-235834P.  
 PR 29-SEP-2000; 2000US-236321P.  
 PR 29-SEP-2000; 2000US-236367P.  
 PR 29-SEP-2000; 2000US-236368P.  
 PR 29-SEP-2000; 2000US-236369P.  
 PR 29-SEP-2000; 2000US-236370P.  
 PR 02-OCT-2000; 2000US-236802P.  
 PR 02-OCT-2000; 2000US-237037P.  
 PR 02-OCT-2000; 2000US-237038P.  
 PR 02-OCT-2000; 2000US-237039P.  
 PR 02-OCT-2000; 2000US-237040P.  
 PR 13-OCT-2000; 2000US-239935P.  
 PR 20-OCT-2000; 2000US-240960P.  
 PR 20-OCT-2000; 2000US-241783P.  
 PR 20-OCT-2000; 2000US-241809P.

PR 01-NOV-2000; 2000US-244617P.  
 PR 17-NOV-2000; 2000US-249299P.  
 PR 08-DEC-2000; 2000US-251856P.  
 PR 08-DEC-2000; 2000US-251868P.  
 PR 08-DEC-2000; 2000US-251869P.  
 XX  
 PA (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 XX  
 XX WPI: 2002-681727/73.  
 DR P-PSDB; ABP66860.  
 XX  
 PT Novel polypeptide useful for diagnosis, prognosis, prevention, and  
 PT treatment of immune, hyperproliferative, renal, respiratory,  
 PT cardiovascular, reproductive, endocrine, gastrointestinal and  
 PT neurological disorders  
 XX  
 PS Claim 1; SEQ ID NO 161; 369pp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (ABV83682-ABV84101) and proteins  
 CC (ABP6710-ABP67129) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 758 BP; 287 A; 164 C; 184 G; 123 T; 0 other;

Query Match 5.4%; Score 47; DB 24; Length 758;  
 Best Local Similarity 54.3%; Pred. No. 0.063;  
 Matches 95; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
 QY 690 GGAGATTCTTGTGTAGTCTATCAGACTACCTTTATCAGCAGCAACTGCTGTTAC 749  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 564 GGAGGTGTATCAGCAGCTCTGGGCCACTGTTTAAACCCCAATAAAGAGCACAGGAT 623  
 QY 750 CAGACACTATACGTCCTCATCATATATGTTAAACAGCAATATCGATAGATAT 809  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 624 ATGACACCCCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 683  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 810 TGAATAATAAATGTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 864  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 684 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 738

RESULT 9  
 ABL33323/C  
 ID ABL33323 standard; DNA; 9741 BP.  
 XX  
 AC ABL33323;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 1296.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antilarisclerotic; antianaemic; cytostatic; neurotropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

RESULT 11	
ABA93725	
ID	ABA93725 standard; cDNA; 3038 BP.
XX	
XX	ABA93725;
XX	
XX	
DT	30-APR-2002 (first entry)
XX	
XX	
DE	Human differentiation/development cDNA clone any2_2bi9.
XX	
KW	Human; foetal brain; foetal kidney; melanoma; testis; anygdala;
KW	gene therapy; chromosome 7p15; ss.
XX	
OS	Homo sapiens.

PR	17-FEB-2000; 2000US-183319P.
PR	16-MAR-2000; 2000US-189862P.
PR	25-MAY-2000; 2000US-207454P.
PR	09-JUN-2000; 2000US-211314P.
PR	18-JUL-2000; 2000US-219007P.
PR	13-DEC-2000; 2000US-255281P.
XX	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX	
XX	Schlegel R, Endege WO, Monahan JE;
XX	WPT; 2001-662795/76.
XX	
PT	Novel isolated nucleic acid molecule associated with, cancerous state of
PT	prostate cells and correlating with presence of prostate cancer, useful
PT	for detecting presence of prostate cancer, stage of prostate cancer -
XX	
PS	Claim 1; Page 11298; 11750pp; English.
XX	
CC	The invention relates to an isolated nucleic acid molecule (I) comprising
CC	a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC	specification or its complement. (I) is useful for:
CC	(a) assessing whether a patient is afflicted with prostate cancer;
CC	(b) monitoring the progression of prostate cancer in a patient;
CC	(c) assessing the efficacy of a test compound to inhibit prostate
CC	cancer in a patient;
CC	(d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC	in a patient;
CC	(e) selecting a composition for inhibiting prostate cancer in a patient;
CC	(f) assessing the prostate cell carcinogenic potential of a compound;
CC	(g) determining whether prostate cancer has metastasized in a patient;
CC	(h) assessing the aggressiveness or indolence of prostate cancer in a
CC	patient;
CC	(I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX	
XX	
SQ	Sequence 488 BP; 229 A; 100 C; 67 G; 92 T; 0 other;
	Query Match            5.2%; Score 45.4; DB 23; Length 488;
	Best Local Similarity 53.0%; Pred. No. 0.14;
	Matches 97; Conservative 0; Mismatches 86; Indels 0; Gaps 0
QY	682 CAGAACTGGAGAATTCCTCGTAGACTCATCAGACACTTTTATCAGCAGACAACGTG 7411
Db	76 CAGAACTATAACATTCAATAGTGTGCTGTGCCTGCTCAAAGTGCTTAGCTCACCTGGATTAOC 135
QY	742 GTCGTTTACAGACACCCTATAACGCTGCTCATCAATATGTTAAAACAGAAATAATCGA 801
Db	136 TCACTGTTAATGCTCAACACAGTAATCAATTAACATTTTGATTAAAAAATAAAAAA 195
QY	802 TAGAATATTGAANAATAATGTTAATAAACACGTGGTGGAAATATGAAAAAATAAAAAA 861
Db	196 AAAAATAA 255
QY	862 AAA 864
Db	256 AAA 258
RESULT 13	
AAL50827	
ID	AAL50827 standard; DNA; 2091 BP.
XX	
AC	AAL50827;
XX	
DT	30-JAN-2003 (first entry)
XX	
DE	Human cancer status prediction method-related DNA sequence #19.
XX	
KW	Human; gene therapy; cancer status prediction; cancer; ds;
KW	cancer malignancy evaluation; drug design; antisense nucleic acid.
XX	
O6	Homo sapiens.
XX	

OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FH	10..909
CDS	/tag= a
FT	/product= BCR2 protein
FF	
FT	
FT	
XX	
PX	WO200055629-A2.
XX	
PD	21-SEP-2000.
PF	
PP	15-MAR-2000; 2000WO-US06952..
PR	
PR	15-MAR-1999; 99US-0258885.
PR	12-NOV-1999; 99US-0439878.
PR	12-NOV-1999; 99US-0440370.
PR	15-NOV-1999; 99US-0440493.
PR	16-NOV-1999; 99US-0440676.
PR	16-NOV-1999; 99US-0440677.
PR	29-NOV-1999; 99US-0450810.
PR	02-DEC-1999; 99US-0453137.
PR	08-MAR-2000; 2000US-0453137.
XX	
PA	(EOSB-) EOS BIOTECHNOLOGY INC.
XX	
PI	Mack D, Gish KC;
XX	
DR	WPI; 2000-638216/61.
DR	P-FSDB; BAB00136.
XX	
PT	Screening drug candidates for their ability to modulate breast cancer
PT	by contacting the drug to a cell expressing an expression profile gene
PT	and determining modulation of expression of the gene
PS	Disclosure; Fig 68; 258pp; English.
XX	
CC	New methods for screening drug candidates are described which
CC	comprise adding a drug candidate to a cell that expresses a protein
CC	selected from BCH1, BCA2, BCJ7, BCN1, BCN5, BCO2, BCO5, BCR2, BCX2
CC	and BCY3 or their fragments and determining the effect of the drug
CC	on the expression of those proteins. Antibodies to breast cancer
CC	genes (specifically BCH1 or its fragment (BCHlp1 or BCHlp2)) are
CC	useful for inhibiting and treating breast cancer in individuals who
CC	are non-responsive to anti-oestrogen and positive for oestrogen
CC	receptor. Compositions comprising BCH1 or a nucleic acid encoding
CC	BCH1 are useful for eliciting an immune response in an individual.
CC	The antibodies are also useful for the diagnosis and prognosis of
CC	breast cancer and for screening compositions which modulate the
CC	breast cancer phenotype. The method allows rapid and simple
CC	detection of lymph node metastases.
XX	
SQ	Sequence 2710 BP; 851 A; 489 C; 525 G; 845 T; 0 other;
	Query Match 5.2%; Score 45.4; DB 21; Length 2710;
	Best Local Similarity 53.0%; Pred. No. 0.23;
	Matches 97; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY	682 CAGAAACTGGAGAATTCCTTGGTAGATCATATCAGACTACTTTTATCACGACGACAACGTG 741
Db	2527 CAGAAACTATAACATTCAATAGTGCTGCTGCACAGTGCTTAGCTCACCTCGATATACC 2586
QY	742 GTCGTTACACACACCTATACAGTGTCCTCATCAATATGTGTAAAACAGAAATTAATCGA 801
Db	2587 TAGATTGTTAAATGCTCAACAGTATCATTAATAAACATTTTGTATTAAAAAATAAAAAA 2646
QY	802 TAGAATATTGAATAAATAAATCTTAAATAAACACTGCTTCAAATATGAAAAAATAAAAAA 861
Db	2647 AAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 2706
QY	862 AAA 864
Db	2707 AAA 2709

Search completed: August 22, 2003, 04:54:50  
Job time : 302 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2003, 01:57:13 ; Search time 72 Seconds

(without alignments)  
5333.380 Million cell updates/sec

Title: US-09-831-142B-1

Perfect score: 870

Sequence: 1 gaattcgacagtcgga.....aaaaaaaaaaaaactcgag 870

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

1: /cgn2.6/ptodata/2/ina/5A.COMB.seq.\*

2: /cgn2.6/ptodata/2/ina/5B.COMB.seq.\*

3: /cgn2.6/ptodata/2/ina/6A.COMB.seq.\*

4: /cgn2.6/ptodata/2/ina/6B.COMB.seq.\*

5: /cgn2.6/ptodata/2/ina/PCTUS.COMB.seq.\*

6: /cgn2.6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	51.6	5.9	7218	1	US-08-232-463-14
2	43	4.9	2665	4	US-08-971-089-5
3	41.6	4.8	1872	4	US-09-801-052-1
4	41.4	4.8	2246	4	US-09-363-708-3
5	41.4	4.8	2246	4	US-09-083-587-3
6	40.2	4.6	6671	1	US-08-280-443-1
7	40.2	4.6	6671	1	US-08-457-459-1
8	40.2	4.6	6671	1	US-08-555-678-1
9	40.2	4.6	6671	5	PCT-US95-02275-1
10	40	4.6	2671	6	5168051-9
11	40	4.6	3275	4	US-09-370-838-151
12	39.6	4.6	1048	4	US-09-489-847-38
13	39.6	4.6	2621	2	US-08-553-619B-8
14	39.4	4.5	903	5	PCT-US95-06408A-21
15	39.2	4.5	2184	3	US-08-955-918C-1
16	39.2	4.5	2184	4	US-08-697-766A-1
17	39	4.5	2806	4	US-09-653-839-9
18	38.8	4.5	2852	3	US-09-027-137-2
19	38.8	4.5	2852	4	US-09-344-441-2
20	38.6	4.4	949	4	US-09-489-847-35
21	38.6	4.4	1052	4	US-09-489-847-23
22	38.6	4.4	1414	4	US-09-501-115-5
23	38.6	4.4	3527	2	US-08-909-965C-7
24	38.6	4.4	6152	3	US-08-973-462-1
25	38.4	4.4	1024	4	US-09-328-475C-50
26	38.4	4.4	1798	4	US-09-797-906-1
27	38.4	4.4	3060	4	US-09-996-243-504

28	38.2	4.4	578	4	US-09-602-877A-95
29	38.2	4.4	1342	4	US-09-489-847-89
30	38.2	4.4	1454	4	US-09-372-422A-19
31	38	4.4	1641	1	US-08-300-903A-8
32	38	4.4	1641	4	US-08-988-197-8
33	37.8	4.3	1582	3	US-08-545-196B-10
34	37.8	4.3	1582	3	US-08-545-196B-12
35	37.8	4.3	2445	6	5215909-9
36	37.6	4.3	1117	3	US-09-247-373B-33
37	37.6	4.3	2320	4	US-09-202-904A-13
38	37.6	4.3	176373	3	US-09-128-155-17
39	37.4	4.3	725	4	US-09-996-243-98
40	37.4	4.3	993	4	US-09-464-535-25
41	37.4	4.3	1114	4	US-09-152-060-41
42	37.4	4.3	2269	4	US-09-394-645-1
43	37.4	4.3	2269	4	US-09-243-560B-1
44	37.4	4.3	2621	2	US-08-553-619B-8
45	37.4	4.3	2674	4	US-09-817-180-1

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: ptz9gt-F1s  
US-08-232-463-14

Query Match 5.9%; Score 51.6; DB 1; Length 7218;

```
Best Local Similarity 4.5%; Pred. No. 0.00011;
Matches 18; Conservative 220; Mismatches 164; Indels 0; Gaps 0;

QY 292 GTGTACAGAAAACAACTCTTCGAGGTTGTGTGAGAAATGCCCGCTAGTTGACT 351
Db 1460 GTTAAAGAGATAGAGAAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRR 1401

QY 352 ACAGACACAGAGAGAGGGAAGTGTGAGAAAATTTTACAGAAAATTTCTAGCAAAATGC 411
Db 1400 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1341

QY 412 CAGGCACTTACATCTATGAGCGTGTGGCTACAGAGGACGCTGATGATNAATGCATCG 471
Db 1340 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1281

QY 472 AAGCACAATTTGGTGCACAGTCCAGGTTGCTCCATATGACGAGATAACAATGCTGTA 531
Db 1280 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1221

QY 532 TGGATGAAGTAAAGTGAATCCATCTGAGACAAATCGAGGATGATATCAAGGACTGTGGC 591
Db 1220 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1161

QY 592 TCTTAGACAGATGTTCACTGATATACGTGACATCAAAACGAGTGTGATCTACGAG 651
Db 1160 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1101

QY 652 ACACAGTACAGCGGCTGAAGACACACACACCGTCCAGAACTGGAG 693
Db 1100 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1059

RESULT 2
US-08-971-089-5
; Sequence 5, Application US/08971089
; Patent No. 6376174
; GENERAL INFORMATION:
; APPLICANT: Puist, Stefan M.
; APPLICANT: Scoles, Daniel R.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING
; TITLE OF INVENTION: SCHWANNOMIN-BINDING-PROTEINS AND PRODUCTS RELATED THERETO
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL & FLORES, LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,089
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/030,987
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ramos, Robert T.
; REGISTRATION NUMBER: 37,915
; REFERENCE/DOCKET NUMBER: P-CE 2862
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)535-9001
; TELEFAX: (619)535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
```

```
; TOPOLOGY: both
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5..2077
US-08-971-089-5

Query Match 4.9%; Score 43; DB 4; Length 2665;
Best Local Similarity 59.3%; Pred. No. 0.019;
Matches 73; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 742 GTCTTACAGACACCTATACGCTGTCTCTCATCAATATGTGTAAACAGAAATATCGA 801
Db 2526 GCCATCTCCAGTGCCTCCAGATGGTACAGCGATATAAATGTATTTCAGAAAAAAA 2585

QY 802 TAGATATTGAAATAAATAATTAATAACACTGTTGCAATATGAAAAAATAAAAA 861
Db 2586 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 2645

QY 862 AAA 864
Db 2646 AAA 2648

RESULT 3
US-09-801-052-1
; Sequence 1, Application US/09801052
; Patent No. 6368842
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE
; FILE REFERENCE: CL001045
; CURRENT APPLICATION NUMBER: US/09/801,052
; CURRENT FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1872
; TYPE: DNA
; ORGANISM: Human
US-09-801-052-1

Query Match 4.8%; Score 41.6; DB 4; Length 1872;
Best Local Similarity 59.2%; Pred. No. 0.04;
Matches 71; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 745 GTTACAGACACCTATACGCTGTCTCTCATCAATATGTGTAAACAGAAATATCGATAG 804
Db 1741 GTTATATAAGTAAATAAATGTTGTTGTTAAATAAAAAAATAAAAAAATAAAAA 1800

QY 805 AATATTGAAATAAATAATGTTAATAACACTGTTGCAATATGAAAAAATAAAAA 864
Db 1801 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 1860

RESULT 4
US-09-363-708-3
; Sequence 3, Application US/09363708
; Patent No. 6399747
; GENERAL INFORMATION:
; APPLICANT: Schmandt, et al.
; TITLE OF INVENTION: NOVEL SHC BINDING PROTEIN
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/363,708  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 01017/34451  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2246 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
DESCRIPTION: /desc = "mouse PAL CDNA"  
US-09-363-708-3

Query Match 4.8%; Score 41.4; DB 4; Length 2246;  
Best Local Similarity 61.7%; Pred. No. 0.05; Indels 0; Gaps 0;  
Matches 66; Conservative 0; Mismatches 41;  
QY 758 TATACCTGCTCCATCAATAATGTGTAACAGAAATATCGATAGATATTGAAATA 817  
DB 2117 TATAGTGTACTTGACCAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 2176  
QY 818 AAATGTTAATAAACACTGGTTGAAATATGAAAAAATAAAAAAATAAAAA 864  
DB 2177 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 2223

RESULT 5  
US-09-083-587-3  
Sequence 3, Application US/09083587  
Patent No. 6492138  
GENERAL INFORMATION:  
APPLICANT: Schmandt, et al.  
TITLE OF INVENTION: NOVEL SHC BINDING PROTEIN  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive/6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,587  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 01017/34451  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2246 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
DESCRIPTION: /desc = "mouse PAL CDNA"  
US-09-083-587-3  
Query Match 4.8%; Score 41.4; DB 4; Length 2246;  
Best Local Similarity 61.7%; Pred. No. 0.05; Indels 0; Gaps 0;  
Matches 66; Conservative 0; Mismatches 41;  
QY 758 TATACCTGCTCCATCAATAATGTGTAACAGAAATATCGATAGATATTGAAATA 817  
DB 2117 TATAGTGTACTTGACCAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 2176  
QY 818 AAATGTTAATAAACACTGGTTGAAATATGAAAAAATAAAAAAATAAAAA 864  
DB 2177 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 2223

RESULT 6  
US-08-280-443-1  
Sequence 1, Application US/08280443  
Patent No. 5643778  
GENERAL INFORMATION:  
APPLICANT: Nishikura, Kazuko  
TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, P.O. Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/280,443  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/197,794  
FILING DATE: 17-FEB-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: WST49AUSA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9206  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6671 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 155..3832  
US-08-280-443-1

Query Match 4.8%; Score 40.2; DB 1; Length 6671;  
Best Local Similarity 64.5%; Pred. No. 0.18; Indels 0; Gaps 0;  
Matches 60; Conservative 0; Mismatches 33;  
QY 772 ATCAATATGTGTAACAGAAATATCGATAGATATTGAAATAATTTAATAAC 831  
DB 6561 AACAGAAATCTGAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 6620

QY 832 ACTGGTTGAATATGAAAAAATAAATAAATAA 864  
 Db 6621 AAAAAAATAAATAAATAAATAAATAAATAA 6653

## RESULT 7

US-08-457-459-1  
 ; Sequence 1, Application US/08457459  
 ; Patent No. 5677428  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nishikura, Kazuko  
 ; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use  
 ; NUMBER OF SEQUENCES: 39  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Howson and Howson  
 ; STREET: Spring House Corporate Cntr, P.O. Box 457  
 ; CITY: Spring House  
 ; STATE: Pennsylvania  
 ; COUNTRY: USA  
 ; ZIP: 19477  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/457,459  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/197,794  
 ; FILING DATE: 17-FEB-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/280,443  
 ; FILING DATE: 25-JUL-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Bak, Mary E.  
 ; REGISTRATION NUMBER: 31,215  
 ; REFERENCE/DOCKET NUMBER: WST49CUSA  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 215-540-9206  
 ; TELEFAX: 215-540-5818  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 6671 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: CDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 155..3832  
 ; US-08-457-459-1

Query Match 4.6%; Score 40.2; DB 1; Length 6671;  
 Best Local Similarity 64.5%; Pred. No. 0.18;  
 Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
 QY 772 ATCAATAATGTGTAATAACAGAAATAATCGATAGAAATTTGCAAAATAAATGTTAATAAC 831  
 Db 6561 AACAGAATCTGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6620

QY 832 ACTGGTTGAATATGAAAAAATAAATAAATAA 864  
 Db 6621 AAAAAAATAAATAAATAAATAAATAAATAA 6653

## RESULT 8

US-08-555-678-1  
 ; Sequence 1, Application US/08555678  
 ; Patent No. 5763174

; GENERAL INFORMATION:  
 ; APPLICANT: Nishikura, Kazuko  
 ; TITLE OF INVENTION: RNA Editing Enzyme and Methods  
 ; TITLE OF INVENTION: of Use Thereof  
 ; NUMBER OF SEQUENCES: 67  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Howson and Howson  
 ; STREET: Spring House Corporate Cntr, P.O. Box 457  
 ; CITY: Spring House  
 ; STATE: Pennsylvania  
 ; COUNTRY: USA  
 ; ZIP: 19477  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/555,678  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/197,794  
 ; FILING DATE: 17-FEB-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/280,443  
 ; FILING DATE: 25-JUL-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/457,459  
 ; FILING DATE: 01-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Bak, Mary E.  
 ; REGISTRATION NUMBER: 31,215  
 ; REFERENCE/DOCKET NUMBER: WST49DUSA  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 215-540-9206  
 ; TELEFAX: 215-540-5818  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 6671 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: CDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 155..3832  
 ; US-08-555-678-1

Query Match 4.6%; Score 40.2; DB 1; Length 6671;  
 Best Local Similarity 64.5%; Pred. No. 0.18;  
 Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
 QY 772 ATCAATAATGTGTAATAACAGAAATAATCGATAGAAATTTGCAAAATAAATGTTAATAAC 831  
 Db 6561 AACAGAATCTGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6620  
 QY 832 ACTGGTTGAATATGAAAAAATAAATAAATAA 864  
 Db 6621 AAAAAAATAAATAAATAAATAAATAAATAA 6653

## RESULT 9

PCT-US95-02275-1  
 ; Sequence 1, Application PC/TUS9502275  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wistar Institute of Anatomy & Biology  
 ; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use  
 ; TITLE OF INVENTION: Thereof  
 ; NUMBER OF SEQUENCES: 39  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Howson and Howson  
 ; STREET: Spring House Corporate Cntr, P.O. Box 457

CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02275  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/280,443  
FILING DATE: 25-JUL-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/197,794  
FILING DATE: 17-FEB-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: WST49BPCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9206  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6671 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 155..3832  
PCT-US95-02275-1

Query Match 4.6%; Score 40.2; DB 5; Length 6671;  
Best Local Similarity 64.5%; Pred. No. 0.16;  
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 772 ATCAATATGTTGAATACAGAAATAATCGATAGAAATATGAAATATAAATGTTAATAAAC 831  
DB 6561 ACACAGATCTGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6620  
QY 832 ACTGGTTGAATATGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 864  
DB 6621 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6653

RESULT 10  
5168051-9  
; Patent No. 5168051  
; APPLICANT: DERVYNCK, RIK M.A.; GOEBDEL, DAVID V.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING TGF-B ITS USES  
; NUMBER OF SEQUENCES: 21  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/389,929  
; FILING DATE: 04-AUG-1989  
; SEQ ID NO: 9  
; LENGTH: 2671  
5168051-9  
Query Match 4.6%; Score 40; DB 5; Length 2671;  
Best Local Similarity 65.9%; Pred. No. 0.14;  
Matches 58; Conservative 0; Mismatches 30; Indels 0; Gaps 0;  
QY 777 TAATGTTGAATATGATAGAAATATGAAATAAATAAATGTTAATAAACACTGG 836  
DB 2573 TAGGTGTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2632  
QY 837 TTGAATATGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 864

Db 2633 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2660

RESULT 11  
US-09-370-838-151  
; Sequence 151, Application US/09370838  
; Patent No. 6444425  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Roadon  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF  
; FILE REFERENCE: 210121.475C1  
; CURRENT APPLICATION NUMBER: US/09/370,838  
; EARLIER FILING DATE: 1999-08-09  
; EARLIER APPLICATION NUMBER: US 09/285,323  
; EARLIER FILING DATE: 1999-04-02  
; NUMBER OF SEQ ID NOS: 289  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 151  
; LENGTH: 3275  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-370-838-151  
Query Match 4.6%; Score 40; DB 4; Length 3275;  
Best Local Similarity 63.5%; Pred. No. 0.15;  
Matches 61; Conservative 0; Mismatches 35; Indels 0; Gaps 0;  
QY 775 AATAATGTTAAACACAGAAATAATCGATAGAAATATGAAATATAAATGTTAATAAACACT 834  
DB 3180 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3239  
QY 835 GGTGTAATATGAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 870  
DB 3240 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3275

RESULT 12  
US-09-489-847-38  
; Sequence 38, Application US/09489847  
; Patent No. 6476195  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al  
; TITLE OF INVENTION: 98 Human Secreted Proteins  
; FILE REFERENCE: PZ031P1  
; CURRENT APPLICATION NUMBER: US/09/489,847  
; CURRENT FILING DATE: 2000-01-24  
; EARLIER APPLICATION NUMBER: PCT/US99/17130  
; EARLIER FILING DATE: 1999-07-29  
; EARLIER APPLICATION NUMBER: 60/094,657  
; EARLIER FILING DATE: 1998-07-30  
; EARLIER APPLICATION NUMBER: 60/095,486  
; EARLIER FILING DATE: 1998-08-05  
; EARLIER APPLICATION NUMBER: 60/096,319  
; EARLIER FILING DATE: 1998-08-12  
; EARLIER APPLICATION NUMBER: 60/095,454  
; EARLIER FILING DATE: 1998-08-06  
; EARLIER APPLICATION NUMBER: 60/095,455  
; EARLIER FILING DATE: 1998-08-06  
; NUMBER OF SEQ ID NOS: 376  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 38  
; LENGTH: 1048  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-489-847-38  
Query Match 4.6%; Score 39.6; DB 4; Length 1048;  
Best Local Similarity 57.1%; Pred. No. 0.11;



```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/955,918C
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/697,766
; FILING DATE: 29-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: MNI-007CPDV2CPA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2184 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 569..1616
US-08-955-918C-1

```

```

Query Match      4.5%; Score 39.2; DB 3; Length 2184;
Best Local Similarity 60.8%; Pred. No. 0.21;
Matches 62; Conservative 1; Mismatches 39; Indels 0; Gaps 0;

QY      763 CGTGCTCCTCATATATGTGTAACACAGAAATATCGATAGATATTTGAAATATAAATG 822
      || ||||| | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2057 CGCGTCTCTGGCMAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2116

QY      823 TTAATAAACACTGGTTGAAATATGAAAAAAAAAAAAAAAAAAAA 864
      || ||||| | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2117 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2159

```

Search completed: August 22, 2003, 06:32:51  
Job time : 75 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: August 22, 2003, 04:53:13 ; Search time 649 Seconds  
(without alignments)  
3013.717 Million cell updates/sec  
Title: US-09-831-142B-1  
Sequence: 1 gaattcgccagctggaa.....aaaaaaaaaaactcgag 870  
Scoring table: IDENTITY\_NUC.  
Gapop 10.0 , Gapext 1.0  
Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	47.4	5.4	9810	12	US-10-311-455-400
C 2	47.2	5.4	3999	14	US-10-198-846-12795
C 3	47	5.4	758	9	US-09-764-833-161
C 4	46.2	5.3	9741	12	US-10-311-455-1296
C 5	45	5.2	17294	12	US-10-311-455-960
C 6	44.8	5.1	805	13	US-10-027-632-29967
C 7	44.8	5.1	805	13	US-10-027-632-29968
C 8	44.8	5.1	6211	12	US-10-311-455-779
C 9	44	5.1	6040	12	US-10-311-455-40
C 10	43.8	5.0	363	10	US-09-960-352-11687
C 11	43.4	5.0	11745	12	US-10-240-453-206
C 12	43.2	5.0	413	11	US-09-918-995-7068
C 13	43.2	5.0	1321	9	US-09-925-299-235
C 14	43.2	5.0	1321	11	US-09-925-299-235
C 15	43.2	5.0	2852	10	US-09-761-640-2
C 16	43.2	5.0	5649	12	US-10-311-455-822

17	43	4.9	1904	12	US-10-137-870-99	Sequence 99, Appl
18	43	4.9	1904	12	US-10-140-018-99	Sequence 99, Appl
19	43	4.9	1904	12	US-10-140-021-99	Sequence 99, Appl
20	43	4.9	1904	12	US-10-140-274-99	Sequence 99, Appl
21	43	4.9	1904	12	US-10-140-471-99	Sequence 99, Appl
22	43	4.9	1904	12	US-10-140-807-99	Sequence 99, Appl
23	43	4.9	1904	12	US-10-140-822-99	Sequence 99, Appl
24	43	4.9	1904	12	US-10-140-924-99	Sequence 99, Appl
25	43	4.9	1904	12	US-10-140-926-99	Sequence 99, Appl
26	43	4.9	1904	12	US-10-141-698-99	Sequence 99, Appl
27	43	4.9	1904	12	US-10-141-702-99	Sequence 99, Appl
28	43	4.9	1904	12	US-10-141-704-99	Sequence 99, Appl
29	43	4.9	1904	12	US-10-142-421-99	Sequence 99, Appl
30	43	4.9	1904	12	US-10-142-432-99	Sequence 99, Appl
31	43	4.9	1904	12	US-10-142-767-99	Sequence 99, Appl
32	43	4.9	1904	12	US-10-143-033-99	Sequence 99, Appl
33	43	4.9	1904	12	US-10-144-994-99	Sequence 99, Appl
34	43	4.9	1904	12	US-10-145-628-99	Sequence 99, Appl
35	43	4.9	1904	12	US-10-145-631-99	Sequence 99, Appl
36	43	4.9	1904	12	US-10-145-633-99	Sequence 99, Appl
37	43	4.9	1904	12	US-10-145-746-99	Sequence 99, Appl
38	43	4.9	1904	12	US-10-145-748-99	Sequence 99, Appl
39	43	4.9	1904	12	US-10-145-823-99	Sequence 99, Appl
40	43	4.9	1904	12	US-10-145-826-99	Sequence 99, Appl
41	43	4.9	1904	12	US-10-145-870-99	Sequence 99, Appl
42	43	4.9	1904	12	US-10-145-876-99	Sequence 99, Appl
43	43	4.9	1904	12	US-10-145-959-99	Sequence 99, Appl
44	43	4.9	1904	12	US-10-146-724-99	Sequence 99, Appl
45	43	4.9	1904	12	US-10-146-725-99	Sequence 99, Appl

ALIGNMENTS

RESULT 1

US-10-311-455-400/c  
; Sequence 400, Application US10311455  
; Publication No. US20030143606A1

; GENERAL INFORMATION:

; APPLICANT: OLEK, Alexander

; APPLICANT: PIEPNBROCK, Christian

; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by De

; FILE REFERENCE: 5013 1014

; CURRENT APPLICATION NUMBER: US/10/311,455

; CURRENT FILING DATE: 2002-12-16

; PRIOR APPLICATION NUMBER: PCT/EP01/07537

; PRIOR FILING DATE: 2001-07-02

; PRIOR APPLICATION NUMBER: DE 10032529.7

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: DE 10043826.1

; PRIOR FILING DATE: 2000-09-01

; NUMBER OF SEQ ID NOS: 2424

; SEQ ID NO 400

; LENGTH: 9810

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-311-455-400

Query Match

Best Local Similarity 5.4%; Score 47.4; DB 12; Length 9810;

Matches 132; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 594 TTAGCCAGAGTGTGACTGCGATTATAGTGGACTCAAAACGAGTGTGATCTACCGAC 653

Db 4058 TTTAACTTACATTTCTACTCTAAATATATATCTTAACACACAACTAAACACCATCTCTAAA 3999

QY 654 ACAGTAGACGAGGTGAGACACACCGTCAGAACTGGAGATCTCTTCTGGTAGATCTAT 713

Db 3998 TCCCAAACTCCGCTAAACAAAACAAAATATTATTACTTACTTCACATATAAATTT 3939



```

; NUMBER OF SEQ ID NOS. 2424
; SEQ ID NO 1296
; LENGTH: 9741
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1296

Query Match: 5.3%; Score 46.2; DB 12; Length 9741;
Best Local Similarity 64.5%; Pred. NO. 0.23;
Matches 69; Conservative 0; Mismatches 38; Indels 0; Gaps 0

Qy 758 TATACTGTGCCATCAATATGTGTAAACACAGAAATATCGATAGAAATATTGAAAAA 817
Db 6660 TATATCTCTTACCTACCAACATTAATCTTAAATATATATATAAACTTTAATCTTTATACAAAA 6601

Qy 818 AAATGTTAATAAACAACCTGGTTGAAATATGAAAAAATAAAAAAAAAA 864
Db 6600 AACATTCATTAACAATAATTTCAATATAATAATAATAATAAAAAA 6554

RESULT 5
US-10-311-455-960/C
; Sequence 960, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16

```

```

; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 960
; LENGTH: 17294
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-960

Query Match
Best Local Similarity 5.2%; Score 45; DB 12; Length 17294;
Matches 60; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 785 AAAACAGAAATGTCAGATATGAAATATAATCTTAATAACACTGCTGTAATA 844
DB 10093 AAAAAAACAATACAAACATTAATAAAAAAATAATAAACCTCTACTAATA 10034
QY 845 TGAATAAAAAAATAAAAACTCGA 869
DB 10033 TCAACAAAAAATAACGCTA 10009

```

```

RESULT 6
US-10-027-632-29967
; Sequence 29967, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29967
; LENGTH: 805
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-29967

```

```

Query Match
Best Local Similarity 5.1%; Score 44.8; DB 13; Length 805;
Matches 64; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 769 CTCATCAATATGTCAGATATGAAATATAATCGATAGATATGAAATATAATGTTAATA 828
DB 301 CCCAGATATTCATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 360
QY 829 AACACTGGTGAATATGAAATATAATAATAATAATAATAATAATAATAATAATAATAATA 864
DB 361 AACAGATGTGGAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 396

```

```

RESULT 7
US-10-027-632-29968
; Sequence 29968, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29968
; LENGTH: 805
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-29968

```

```

Query Match
Best Local Similarity 5.1%; Score 44.8; DB 13; Length 805;
Matches 64; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 769 CTCATCAATATGTCAGATATGAAATATAATCGATAGATATGAAATATAATGTTAATA 828
DB 301 CCCAGATATTCATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 360
QY 829 AACACTGGTGAATATGAAATATAATAATAATAATAATAATAATAATAATAATAATAATA 864
DB 361 AACAGATGTGGAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 396

```

```

RESULT 8
US-10-311-455-779/c
; Sequence 779, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by De
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 779
; LENGTH: 6211
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-779

```

```

; SEQ ID NO 11687
; LENGTH: 363
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 50-LIB3058-012-Q1-K1-E10
US-09-960-352-11687

Query Match          5.0%; Score 43.8; DB 10; Length 363;
Best Local Similarity 56.6%; Pred. No. 0.18;
Matches 81; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 722 TTTTATCAGCAGCAACTGCTGCTACCCAGACACCTATAACGCTGCTCCTCATCAATAATG 781
Db 171 TTTTCTCCCTTAAATTCAGCTGCTTTCCAGACAGCAGCCACTTTTACTTTTAAACAATA 112
QY 782 TGTAAACAGAAATTAATCGATPAGATATTGAAATAAATAATGTTAATAAACACATGCTGAA 841
Db 111 TAAAAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 52
QY 842 ATATGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 864
Db 51 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 29

RESULT 11
US-10-240-453-206/c
; Sequence 206, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: By Means of Assessing the Methylation Status of Genes Associated
; TITLE OF INVENTION: With DNA Transcription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 206
; LENGTH: 11745
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (9105)
US-10-240-453-206

Query Match          5.0%; Score 43.4; DB 12; Length 11745;
Best Local Similarity 53.9%; Pred. No. 1.3;
Matches 89; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 697 TCTTCTGGTAGATCTATCAGACTACTTTTATCAGCAGGACAACTGGTGTACCGACAC 756
Db 6549 TATTCCTAAATTACTATAAATAATTAATAATTTAACTTTAAATAATCAATTAATAAATAAAT 6490
QY 757 CTATAAGCTGCTCTCATTAATGTTGTAACAGAAATAATCCGATAGATATTGAAAT 816
Db 6489 CTACAAAAACAATAAATAAATAAAGAGCTCTATAACAACTCACCAGAAATATCGAAATTTAAAAA 6430

```

Qy	861
Db	6385

RESULT 12

```

US-09-918-995-7068
; Sequence 7068, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7068
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(413)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-7068

```

Query Match 5.0%; Score 43.2; DB 11; Length 413;  
Best Local Similarity 57.7%; Pred. No. 0.28;  
Matches 75; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY	734	GACAACTGGTGGTTACCA	GACACACCTATTAACG	TGTCCTCATCAATAATGTG	TAAACAGAA	793
Db	101	GATGACTTGACACGCCCT	CCCCCAATAAACTGC	GCCTCTTGGAATGGC	NNAAAAAAA	160

[illegible]

QY	854	AAAAAAAAAA	863
Db	221	AAAAAAAAAA	230

RESULT 13

```

US-09-925-299-235
? Patent 235, Application US/09925299
? Sequence No. US20020055627A1
? GENERAL INFORMATION:
? APPLICANT: Rosen et al.
? TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
? FILE REFERENCE: PA102
? CURRENT APPLICATION NUMBER: US/09/925,299
? CURRENT FILING DATE: 2001-08-10
? PRIOR APPLICATION NUMBER: PCT/US00/05683
? PRIOR FILING DATE: 2000-03-08
? PRIOR APPLICATION NUMBER: 60/124,270
? PRIOR FILING DATE: 1999-03-12
? NUMBER OF SEQ ID NOS: 1556
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 235
? LENGTH: 1321
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1313)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc_feature
? LOCATION: (1320)

```

; OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-299-235

	Query Match	5.0%;	Score 43.2;	DB 9;	Length 1321;
	Best Local Similarity	63.5%;	Pred. No. 0.5;		
	Matches 66;	Conservative	Mismatches 38;	Indels 0;	Gaps 0;
Qy	761	AACTGTCTCTCAATAATGTGTAAACAGCAAAATAATCGTAGAATATTGAAATAATAAA	820		
Db	1205	AAAGTATCTCTCTCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	1264		
Qy	821	TGTTTAATAACACTGGTGTGAAATATGAAAAAATAAAAAAATAAAAAA	864		
Db	1265	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	1308		

RESULT 14

```

US-09-925-299-235
? Sequence 235, Application US/09925299
? Publication No. US20030040617A9
? GENERAL INFORMATION:
? APPLICANT: Rosen et al.
? TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
? FILE REFERENCE: PAl02
? CURRENT APPLICATION NUMBER: US/09/925,299
? CURRENT FILING DATE: 2001-08-10
? PRIOR APPLICATION NUMBER: PCT/US00/05683
? PRIOR FILING DATE: 2000-03-08
? PRIOR APPLICATION NUMBER: 60/124,270
? PRIOR FILING DATE: 1999-03-12
? NUMBER OF SEQ ID NOS: 1556
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 235
? LENGTH: 1321
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1313)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc_feature
? LOCATION: (1320)
? OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-235

```

[illegible]

## RESULT 15

```

US-09-761-640-2
; Sequence 2, Application US/09761640
; Patent No. US20020137042A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CLO00964-CIP
; CURRENT APPLICATION NUMBER: US/09/761,640
; CURRENT FILING DATE: 2001-01-18
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

```

Search completed: August 22, 2003, 06:43:51  
Job time : 652 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2003, 01:42:03 ; Search time 2303 Seconds  
(without alignments)  
9181.456 Million cell updates/sec

Title: US-09-831-142B-1

Perfect score: 870

Sequence: 1 gaattcgcaagtcgagaa.....aaaaaaaaaaaaaacctcgag 870

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estau.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_htc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_htc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_nam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rod.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50.8	5.8	918	9	AL544003
2	50.2	5.8	683	12	BI524329
3	49.8	5.7	206	14	C89835
4	49.8	5.7	317	14	CB708528

5	49.4	5.7	310	9	AA230648
6	49.4	5.7	811	12	BI825633
7	48.2	5.5	385	10	BE048135
8	48	5.5	792	28	AZ196391
9	47.6	5.5	338	14	CA338213
10	47.6	5.5	346	12	BI500307
11	47.6	5.5	815	13	BU566867
12	47.4	5.4	349	13	BU589755
13	47.4	5.4	430	13	BU564850
14	47.4	5.4	438	14	CA337338
15	47.4	5.4	1065	29	CNS074RT
16	47.4	5.4	1121	29	CNS01623
17	47	5.4	408	14	CA779772
18	47	5.4	1102	13	EX381089
19	46.8	5.4	289	12	BM184688
20	46.8	5.4	978	9	AL544049
21	46.8	5.4	1041	14	CD051043
22	46.6	5.4	211	13	BO667088
23	46.6	5.4	325	14	C93882
24	46.6	5.4	955	13	BU829015
25	46.6	5.4	1108	29	CNS00710
26	46.4	5.3	170	12	BM495105
27	46.4	5.3	515	9	AI399560
28	46.4	5.3	1115	13	BU531671
29	46.2	5.3	445	12	BI745651
30	46.2	5.3	477	12	BM309523
31	46.2	5.3	935	13	EX328353
32	46.2	5.3	956	29	CNS014V0
33	46.2	5.3	975	29	CNS03DC6
34	46	5.3	236	13	BQ937940
35	46	5.3	375	14	CA335987
36	46	5.3	553	9	AL633439
37	46	5.3	669	11	AY058474
38	46	5.3	808	14	CB318829
39	46	5.3	1264	13	BQ940146
40	45.8	5.3	139	12	BI814718
41	45.8	5.3	213	9	AW423994
42	45.8	5.3	225	13	BU842591
43	45.8	5.3	371	14	CA301774
44	45.8	5.3	898	11	EC017007
45	45.8	5.3	935	13	EX374628

#### ALIGNMENTS

RESULT 1  
AL544003/c  
LOCUS  
DEFINITION AL544003 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CSODI004YA01 5-PRIME, mRNA sequence.  
ACCESSION AL544003  
VERSION EST.  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

918 bp mRNA linear EST 31-MAY-2003  
AL544003 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CSODI004YA01 5-PRIME, mRNA sequence.

AL544003  
GI:31265848

EST.  
Homo sapiens (human)

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 918)

Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

Full-length cDNA libraries and normalization

Unpublished

On Feb 15, 2001 this sequence version replaced gi:12876482.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by life technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 8347.r For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgf-bin/cluster.cgi?seq=CSODI004AA01Q1&cluster=8347.r. Contact :

Feng Liang Email : fliang@lifetech.com URL :

```

FEATURES
  source
    1. .683
    Location/Qualifiers
      org_qiury:sequence:500..67.
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:5201637"
      /lab_host="DH10B"
      /clone_lib="NIH_MGC_122"

```

CONTACT: Robert Strausberg, Ph.D.  
Email: cxa@nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L7AM1505 row: d column: 22  
High quality sequence stop: 87.

```

Query Match          5.7%; Score 49.8; DB 14; Length 206;
Best local similarity 59.6%; Pred.No. 8.8e+02;
Matches 84; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

724 TTATCAGCAGGACAACTGTCGTTCACAGACACCTTAACGTCCTCATCAATAATGTG 783
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db    28 TTATCTTTTAAACAATGGGATTTTAAAAACTCTTCAATGTTATTAATATTTTAAATG 87
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

QY 784 TAAACAGAAATATTCGATAGATATTTGAAATAAATGTTAATAAACACTGGTGAAT 843
Db 88 CTAAACCTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 147
QY 844 ATCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 864
Db 148 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 168

RESULT 4
LOCUS CB708528
DEFINITION AMGNUG:SRCS1-00022-G3-A srcs1 (10883) Rattus norvegicus cDNA clone
ACCESSION CB708528
VERSION CB708528.1 GI:29765676
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 317)
AUTHORS Amgen EST Program.
TITLE Amgen Rat EST Program
JOURNAL Unpublished
COMMENT Contact: Dan Fitzpatrick
Amgen, Inc
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00022 row: g column: 3.
FEATURES
source
1..317
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="srcs1-00022-g3"
/tissue_type="Cornea"
/clone_lib="srcs1 (10883)"
/notes="Vector: PCR4-TOPO; Site_1: EcoRI; Site_2: EcoRI;
rat corneas"
BASE COUNT 205 a 27 c 28 g 57 t
ORIGIN
Query Match 5.7%; Score 49.8; DB 14; Length 317;
Best Local Similarity 54.0%; Pred. No. 7.1e-02;
Matches 102; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 676 CACCCTCAGAACTGGAGAAATTCCTCTGGTAGATCTATCAGACACACTTTTATCAGCAGGA 735
Db 83 CAGCCACAGAGTCAGTATATTTTCAGAGAAAGTTATCCACTCAATTTTCTGAATGA 142
QY 736 CAACCTGGCTGTACAGACACCTATACCTGTCCTCCTCAATCAATATGTTAAACAGAAAT 795
Db 143 TAATTAACTTTCTGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 202
QY 796 ATCCGATAGATATTTGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 855
Db 203 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 262
QY 856 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 864
Db 263 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA

RESULT 5
LOCUS AA230648
DEFINITION mw06d10.r1 Soares mouse 3NME12.5 Mus musculus cDNA clone
IMAGE:669907 5' similar to gb:X51438 Mouse mRNA for vimentin (MOUSE
);, mRNA sequence.
ACCESSION AA230648

```

```

VERSION AA230648.1 GI:1852963
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 310)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:409611
Putative full length read
vector to vector length is 733
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 282.
FEATURES
source
1..310
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:669907"
/sex="unknown"
/tissue_type="fetus"
/dev_stage="12.5dpc total fetus"
/lab_host="DH10B"
/clone_lib="Soares mouse 3NME12.5"
/notes="Organ: whole fetus; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5' TGTTACCAATCTGAAGTGGAGCGCGCCCTATTTTATTTTATTTTATTTT
3'], on total mouse RNA [provided by Minoru Ko, Wayne
State Univ.]; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 179 a 33 c 36 g 62 t
ORIGIN
Query Match 5.7%; Score 49.4; DB 9; Length 310;
Best Local Similarity 55.6%; Pred. No. 8.2e-02;
Matches 95; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 694 AATCTCTCTGGTAGATCTATCAGACACTATTTATCAGCAGCAGCACTGCTGCTACCAGA 753
Db 116 AAGTATTTTGTAATACCATTAATAAACTGCTTTTCCAGTAAATATCTGCACCACTTGTTA 175
QY 754 CACCTATACGTCGCTCATCAATAATGTGTAAACACAGAAATATCGATAGATATGAA 813
Db 176 CTGCTTCATTAATCTTCAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 235
QY 814 AATAAAATGTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 864
Db 236 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 286
RESULT 6
LOCUS BT825633

```

BT825633 811 bp mRNA linear EST 04-OCT-2001



```

DEFINITION 603072606f1 NIH_MGC_l19 Homo sapiens cDNA clone IMAGE:5164560 5',
            mRNA sequence.
ACCESSION  BI825633
VERSION     BI825633.1  GI:15937183
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens

REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (Bases 1 to 811)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM1408 row: 1 column: 01
            High quality sequence stop: 276.
            Location/Qualifiers
                1..811
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5164560"
                /tissue_type="medulla"
                /lab_host="DH10B"
                /clone_lib="NIH_MGC_l19"
                /notes="Organ: brain; Vector: pCMV-SPORT6; Site:1: NotI;
                Site:2: EcoRV (destroyed); RNA source normal medulla from
                anonymous male age 27. Library is oligo-dT primed and
                directionally cloned (EcoRV site is destroyed upon
                cloning). Average insert size 1.3 kb, insert size range
                0.9-3 kb. Library is normalized and enriched for
                full-length clones and was constructed by C. Gruber
                (Invitrogen). Research Genetics tracking code 013. Note:
                This is a NIH-MGC Library."
            BASE COUNT  319 a 182 c 163 g 147 t
            ORIGIN
            Query Match 5.7%; Score 49.4; DB 12; Length 811;
            Best Local Similarity 58.5%; Pred. No. 5e+02;
            Matches 86; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

            QY 718 CTACTTTTATCAGCAGCACTGGTTCGTATACACACACCTATAAGCTGTCTCATCAAT 777
            Db 119 CTACTTTTATCAGCAGCACCAGCAATGATGCTTAAATACATAAGATCTCAAGAAAAA 178

            QY 778 AATGTGTAACAGAAATATCGATAGAAATTTGAAATTAATGTTAATAACACACTGGT 837
            Db 179 AAAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 238

            QY 838 TGAATATGAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 864
            Db 239 AAAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 265

            RESULT 7
            LOCUS BE048135
            DEFINITION tz46c11.y1 NCI_CGAP_Brn52 Homo sapiens cDNA clone IMAGE:2291636 5',
            mRNA sequence.
            ACCESSION BE048135
            VERSION BE048135
            KEYWORDS EST.
            SOURCE Homo sapiens (human)
            ORGANISM Homo sapiens

            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (Bases 1 to 385)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute Of Neurological
        Disorders and Stroke, Brain Tumor Genome Anatomy Project
        (CGAP/BTGAAP), Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
        Email: cgapbs-remail.nih.gov
        Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
        R. Emmert-Buck, M.D., Ph.D.
        cDNA Library Preparation: Life Technologies, Inc.
        DNA Sequencing by: Greg Leannon, Ph.D.
        cDNA Library Arrayed by: Washington University Genome Sequencing Center
        Clone distribution: NCI-CGAP clone distribution information can be
        found through the I.M.A.G.E. Consortium/LLNL, send email to:
        info@image.llnl.gov
        Insert Length: 437 Std Error: 0.00
        Seq primer: -40RP from Gibco
        High quality sequence stop: 369.
        Location/Qualifiers
            1..385
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:2291636"
            /tissue_type="tumor, 5 pooled (see description)"
            /lab_host="DH10B"
            /clone_lib="NCI_CGAP_Brn52"
            /notes="Organ: brain; Vector: pCMV-SPORT6; Site:1: SalI;
            Site:2: NotI; This library represents the normalized
            version of NCI_CGAP_Brn35. Cloned unidirectionally.
            Primer: oligo dT. Average insert size 1.19 kb. Tumor
            types include: meningioma, oligodendroglioma, astrocytoma
            (grade II), medulloblastoma, astrocytoma (grade IV).
            Constructed by Life Technologies."
            BASE COUNT  210 a 25 c 58 g 92 t
            ORIGIN
            Query Match 5.5%; Score 48.2; DB 10; Length 385;
            Best Local Similarity 51.6%; Pred. No. 1.1e+03;
            Matches 110; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

            QY 652 ACACAGTAGAGGCGTGAAGACACACCGTCAGAAATCTCTCTGTAGATCT 711
            Db 105 ATACATAGAGGTGGCTCAAAAGTGTTCAGACACCTCGTATTCCTTAATGCGGAGT 164

            QY 712 ATCAGACTACTTTTATCAGCAGGACAACTGGTCGTTTACCAGACACCTATAAGCTGTCTC 771
            Db 165 GACTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 224

            QY 772 ATCAATATGTTAAACAGAAATTAATCGATAGAAATTTGAAATTAATTTATTAATAC 831
            Db 225 TTATTAATTTTAAATTTTGAAGATAAAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTT 284

            QY 832 ACTGCTTGAATATGAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 864
            Db 285 AAAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 317

            RESULT 8
            LOCUS AZ196391/c
            DEFINITION SP_1031_B2_G05_T7A Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone plate-1031 Col-10 Row-N, genomic survey sequence.
            ACCESSION AZ196391
            VERSION AZ196391.1  GI:8390214
            KEYWORDS GSS.
            SOURCE Strongylocentrotus purpuratus
            ORGANISM Strongylocentrotus purpuratus

            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;

```

Strongylocentrotidae; Strongylocentrotus.  
1 (bases 1 to 792)  
Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L.  
A sea urchin genome project: Sequence scan, virtual map, and additional resources  
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)  
20402566  
10920195  
Contact: Cameron, R.A., Davidson, E.H., Hood, L.  
Division of Biology 156-29  
California Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: acameron@caltech.edu  
Plate: 1031 row: N column: 10  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 792.  
Location/Qualifiers  
1..792  
/organism="Strongylocentrotus purpuratus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7668"  
/clone="Plate=1031 Col=10 Row=N"  
/clone\_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"  
/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"

BASE COUNT 104 a 74 c 158 g 406 t 50 others  
ORIGIN  
Query Match 5.5%; Score 48; DB 28; Length 792;  
Best Local Similarity 50.0%; Pred. No. 8e+02;  
Matches 111; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
QY 643 ATCTACACAGACAGTAGTACGAGCGTGTGAGACACACCGTCAGAACTGGAGATTCCTCT 702  
Db 721 AACACACACAAACAAACAAACAAACAAACACACACACACACACACACACACACACCC 662  
QY 703 GGTAGATCTATCAGACTACTTTTATCAGCAGGACAACTGGTGTACACAGACACCTATAA 762  
Db 661 AACACACACAAACAAACAAACAAACAAACAAACACACACACACACACACACAAACAA 602  
QY 763 CGTGCTCTATCAATATGTGTAAACAGAAATATATCGATAGATATTTGAAATATAATG 822  
Db 601 AAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 542  
QY 823 TTAATAAACACTGCTTGAATATGAAATATGAAATATGAAATATGAAATATGAAATAT 864  
Db 541 NAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 500

RESULT 9  
CA338213  
LOCUS  
DEFINITION  
NISC\_lw12f01.y1 COGENE 4PAL Homo sapiens cDNA clone IMAGE:5610001  
5', mRNA sequence.  
CA338213  
VERSION  
CA338213.1 GI:24556311  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 338)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished

Strongylocentrotidae; Strongylocentrotus.  
1 (bases 1 to 792)  
Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L.  
A sea urchin genome project: Sequence scan, virtual map, and additional resources  
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)  
20402566  
10920195  
Contact: Cameron, R.A., Davidson, E.H., Hood, L.  
Division of Biology 156-29  
California Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: acameron@caltech.edu  
Plate: 1031 row: N column: 10  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 792.  
Location/Qualifiers  
1..792  
/organism="Strongylocentrotus purpuratus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7668"  
/clone="Plate=1031 Col=10 Row=N"  
/clone\_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"  
/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"

BASE COUNT 104 a 74 c 158 g 406 t 50 others  
ORIGIN  
Query Match 5.5%; Score 48; DB 28; Length 792;  
Best Local Similarity 50.0%; Pred. No. 8e+02;  
Matches 111; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
QY 643 ATCTACACAGACAGTAGTACGAGCGTGTGAGACACACCGTCAGAACTGGAGATTCCTCT 702  
Db 721 AACACACACAAACAAACAAACAAACAAACACACACACACACACACACACACACACCC 662  
QY 703 GGTAGATCTATCAGACTACTTTTATCAGCAGGACAACTGGTGTACACAGACACCTATAA 762  
Db 661 AACACACACAAACAAACAAACAAACAAACAAACACACACACACACACACACAAACAA 602  
QY 763 CGTGCTCTATCAATATGTGTAAACAGAAATATATCGATAGATATTTGAAATATAATG 822  
Db 601 AAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 542  
QY 823 TTAATAAACACTGCTTGAATATGAAATATGAAATATGAAATATGAAATATGAAATAT 864  
Db 541 NAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 500

RESULT 9  
CA338213  
LOCUS  
DEFINITION  
NISC\_lw12f01.y1 COGENE 4PAL Homo sapiens cDNA clone IMAGE:5610001  
5', mRNA sequence.  
CA338213  
VERSION  
CA338213.1 GI:24556311  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 338)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
CDNA Library Preparation:  
DNA Sequencing by: The I.M.A.G.E. Consortium/LLNL  
Sequencing Center (NISC)

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
info@image.llnl.gov

Plate: LLAM12415 row: L column: 2  
Seq primer: M13Rpl reverse primer (ABI).  
Location/Qualifiers  
1..338

FEATURES  
source

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5610001"  
/tissue\_type="1st pharyngeal arch (pooled)"  
/dev\_stage="embryo, 4 weeks postconception"  
/lab\_host="DH10B"  
/clone\_lib="COGENE 4PAL"  
/note="Vector: pAMP1; cDNA primed using oligo-dT primer, directionally cloned into UDG sites of pAMP1. Size selected for insert sizes ranging from 0.2-1.2 kb. Normalized to Cot10. Primary library, non-amplified. Library constructed by M. Lovett. For more information on this library, please contact R. Tidwell (Washington University) or visit the COGENE website at <http://hg.wustl.edu/COGENE/>."

BASE COUNT 151 a 41 c 43 g 102 t 1 others  
ORIGIN

Query Match 5.5%; Score 47.6; DB 14; Length 338;  
Best Local Similarity 53.6%; Pred. No. 1.4e-03;

Matches 98; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 682 CAGAAACTGGAGATTCCTTCTGTAGATCTATCAGACTACTTTTATCAGCAGCAGCACTG 741

Db 156 CAGAACTATACATCAATAGTGTCTGTCAAGTGTCTAGCTCCTCTAGCTCCTGGATATAC 215

QY 742 GTCTTACACACACCTATACGTCCTCATCAATAATGTGTAAACAGAAATATCGA 801

Db 216 TACATTGTAAATGCTAAACAGTAACTATCAATTAACATTTTGTATTANAAAAA 275

QY 802 TAGAATATTGAAATATAAATCTTAATAAACACTGGTTGAATATGAAAAA 861

Db 276 AAAAAA 335

QY 862 AAA 864

Db 336 AAA 338

## RESULT 10

## BI500307

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.

## 1 (bases 1 to 346)

## McCart, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelis, V., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

REFERENCE	mammalia; eutheria; primates; carnivorae; hominidae; homo.
AUTHORS	1 (bases 1 to 349)
TITLE	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cyabps-re@mail.nih.gov">cyabps-re@mail.nih.gov</a> Tissue Procurement: NCI DNA Library Preparation: Michael Brownstein Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>
	plate: trfcm250n row: 4 column: 08



found through the I.M.A.G.E. Consortium/LINL at:

info@image.lnl.gov

Plate: LHAM2412 row: B column: 24

Seq primer: M13RPI reverse primer (ABI).

Location/Qualifiers

## FEATURES

source

1. 438

/organism="Homo sapiens"

/mol\_type="rRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:560831"

/tissue\_type="posterior rhombomeres 5-8 (4 pooled)"

/dev\_stage="embryo, 4 weeks postconception"

/lab\_host="DH10B"

/clone\_lib="COGENE PR (4EPR)"

/notes="Vector: PAMPL; cDNA primed using oligo-AT primer,

directionally cloned into UDG sites of pAMP1. Size

selected for insert sizes ranging from 0.2-2.0 kb.

Normalized to Cot10. Primary library, non-amplified.

Library constructed by M. Lovett. For more information on

this library, please contact R. Tidwell (Washington

University) or visit the COGENE website at

http://hg.wustl.edu/COGENE/."

161 a 109 c 86 g 80 t 2 others

BASE COUNT

ORIGIN

Query Match 5.4%; Score 47.4; DB 14; Length 438;

Best Local Similarity 57.1%; Pred. No. 1.3e+03;

Matches 84; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 718 CTACTTTTATCAGCAGCAGCACTGTCGTACACAGCACTTAACGTCTCTCTAT 777

Db 287 CCACATATTTCTCCAGTCTCTCCGCCACCCAGCATCAATAAGTGTCAACNGA 346

QY 778 AATGTGTAACACAGAAATATCGATAGATATTCGAAATAAATGTAATAACACTGGT 837

Db 347 ANAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 406

QY 838 TGAATATGAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 864

Db 407 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 433

## RESULT 15

CNS074RT

LOCUS

clone BA0AB030C03 of library BA0AB from strain CLIB 210 of

Kluyveromyces lactis, genomic survey sequence.

AL429183

AL429183.1 GI:12212377

GSS

Kluyveromyces lactis

Kluyveromyces lactis

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

1 (bases 1 to 1065)

Souciot J.L., Aigle, M., Artiguenave, P., Blandin, G.,

Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,

Ge-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Lorente, B.,

Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,

Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,

Wincker, P. and Weissbach, J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

20584711

11152876

2 (bases 1 to 1065)

Bolotin-Fukuhara, M., Toffano-Nioche, C., Artiguenave, F.,

Duchateau-Nguyen, G., Lemaire, M., Marmelisse, R., Montrocher, R.,

Robert, C., Terrier, M., Wincker, P. and Wesolowski-Louvel, M.

Genomic exploration of the hemiascomycetous yeasts: 11.

Kluyveromyces lactis

FEBS Lett. 487 (1), 66-70 (2000)

JOURNAL

TITLE

MEDLINE 20584721

PUBMED 11152886

REFERENCE 3 (bases 1 to 1065)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL

Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,

2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :

seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen

yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces

exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,

Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces

lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia

angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,

Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to

5 kb were prepared and both extremities were sequenced. See

keywords for description of this sequence and for the sequence of

the other extremity of this insert.

## FEATURES

source

1. 1065

/organism="Kluyveromyces lactis"

/mol\_type="genomic DNA"

/strain="CLIB 210"

/variety="lactis"

/db\_xref="taxon:28985"

/clone="BA0AB030C03"

/clone\_lib="BA0AB"

424 a 125 c 104 g 275 t 137 others

BASE COUNT

ORIGIN

Query Match 5.4%; Score 47.4; DB 29; Length 1065;

Best Local Similarity 49.0%; Pred. No. 8.4e+02;

Matches 76; Conservative 13; Mismatches 66; Indels 0; Gaps 0;

QY 710 CTATCAGACTACTTTTATCATCAGCAGCACTGTCGTACACACACTTAAACGTGTC 769

Db 728 CMWTAATAWATATATKANNACNNATWHTAAATAATTTWATATAAATAATWTTT 787

QY 770 TCATCAATAATGTGTAAACAGAAATAATCGATAGATAATTTGAAATAAATAATGTAATA 829

Db 788 TTATCAAAAWNTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 847

QY 830 ACACGTGTTGAATATCAATAAATAAATAAATAAATAAATAAATAAATAAATAA 864

Db 848 AAAATAAKWAATTTATATAAATAAATAAATAAATAAATAAATAAATAAATAA 882

Search completed: August 22, 2003, 06:31:32

Job time : 2311 secs